

**This Page Is Inserted by IFW Operations
and is not a part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/12, C07K 14/475, A61K 38/18, C07K 16/22, C12Q 1/68		A2	(11) International Publication Number: WO 99/38976 (43) International Publication Date: 5 August 1999 (05.08.99)
(21) International Application Number: PCT/US99/00654 (22) International Filing Date: 12 January 1999 (12.01.99) (30) Priority Data: 09/015,412 29 January 1998 (29.01.98) US (63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application US 09/015,412 (CIP) Filed on 29 January 1998 (29.01.98) (71) Applicant (for all designated States except US): INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #12, Mountain View, CA 94040 (US). LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). CORLEY, Neil, C. [US/US]; 1240 Dale Avenue #30, Mountain View, CA 94040 (US). BAUGHN, Mariah [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). GUEGLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US).			TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). (74) Agents: BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US). (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>Without international search report and to be republished upon receipt of that report.</i>
(54) Title: HUMAN GROWTH FACTOR HOMOLOGS (57) Abstract The invention provides human growth factor homologs (HGFH) and polynucleotides which identify and encode HGFH. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for treating or preventing disorders associated with expression of HGFH.			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

HUMAN GROWTH FACTOR HOMOLOGS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of human growth
5 factor homologs, and to the use of these sequences in the diagnosis, treatment, and
prevention of cell proliferative and developmental disorders.

BACKGROUND OF THE INVENTION

Intercellular communication is essential for the development and survival of
10 multicellular organisms. Communication is achieved through the secretion of proteins by
signaling cells and the internalization of these proteins by their target cells. Growth
factors are an example of secreted proteins that mediate communication between signaling
and target cells. Inside the signaling cell, growth factors are synthesized and transported
through the secretory pathway. Entry into the secretory pathway is mediated by the signal
15 peptide sequence, a protein sorting motif at the N-terminus of most secreted proteins.
Within the secretory pathway, the signal sequence is proteolytically removed from its
cognate growth factor. Most growth factors also undergo further post-translational
modifications within the secretory pathway. These modifications can include
glycosylation, phosphorylation, and intramolecular disulfide bond formation. Following
20 their secretion into the extracellular space, some growth factors oligomerize or associate
with extracellular matrix components. The secreted growth factors bind to specific
receptors on the surfaces of their target cells, and the bound receptors trigger second
messenger signal transduction pathways. These signal transduction pathways elicit
specific cellular responses in the target cells. These responses can include the modulation
25 of gene expression and the stimulation or inhibition of cell division, cell differentiation,
and cell motility.

Most growth factors are local mediators that act on cells in the immediate
environment. Such local activity is maintained by physical proximity of the signaling cell
to its target cell, sequestration of the growth factor by extracellular matrix components,
30 internalization and degradation of the growth factor by the target cell, and exclusion of the
growth factor from circulation.

Growth factors fall into three broad and overlapping classes. The first and broadest class includes the large polypeptide growth factors, which are wide-ranging in their effects. These factors include epidermal growth factor (EGF), fibroblast growth factor (FGF), transforming growth factor- β (TGF- β), insulin-like growth factor (IGF), nerve growth factor (NGF), and platelet-derived growth factor (PDGF), each defining a family of numerous related factors. The large polypeptide growth factors generally act as mitogens on diverse cell types to stimulate wound healing, bone synthesis and remodeling, extracellular matrix synthesis, and proliferation of epithelial, epidermal, and connective tissues. Some members of the TGF- β , EGF, and FGF families also function as inductive signals in the differentiation of embryonic tissue. However, some of the large polypeptide growth factors carry out specific functions on a restricted set of target tissues. For example, mouse growth/differentiation factor 9 (GDF-9) is a TGF- β family member that is expressed solely in the ovary. (McPherron, A. C. and Lee, S.-J. (1993) J. Biol. Chem. 268:3444-3449.) NGF functions specifically as a neurotrophic factor, promoting neuronal growth and differentiation.

Large polypeptide growth factors and the signal transduction pathways they trigger are often conserved among eukaryotes ranging from nematodes, fruit flies, yeast, and mammals. Prokaryotes also produce "growth factors" required for morphogenetic and life cycle transitions such as sporulation. (Albertini, A. M. et al. (1987) J. Bacteriol. 169:1480-1484.) Recently, a yeast protein similar to bacterial sporulation outgrowth factor B has been identified. (Johnston, M. et al. (1994) Science 265:2077-2082.)

The second class of growth factors includes the hematopoietic growth factors, which have a narrow target specificity. These factors stimulate the proliferation and differentiation of blood cells such as B-lymphocytes, T-lymphocytes, erythrocytes, platelets, eosinophils, basophils, neutrophils, macrophages, and their stem cell precursors. These factors include the colony-stimulating factors (e.g., G-CSF, M-CSF, GM-CSF, and CSF1-3), erythropoietin, and the cytokines. Cytokines are specialized hematopoietic factors secreted by cells of the immune system in response to external insults, such as tissue damage and viral or microbial infection. Cytokines function in tissue repair, inflammation, and modulation of the immune response. Cytokines include the interleukins, IL1-IL13; the interferons, IFN- α , - β , and - γ ; and the tumor necrosis factors, TNF- α and - β .

The third class of growth factors includes the small peptide factors, which primarily function as hormones in the regulation of highly specialized processes other than cellular proliferation. These factors, which are typically less than 20 amino acids in length, are generated by the proteolytic processing of larger precursor proteins. Some of these factors include bombesin, vasopressin, oxytocin, endothelin, transferrin, angiotensin II, vasoactive intestinal peptide, bradykinin, and related peptides. (Pimentel, E. (1994) Handbook of Growth Factors, CRC Press, Ann Arbor, MI; McKay, I. and Leigh, I., eds. (1993) Growth Factors: A Practical Approach, Oxford University Press, New York, NY; and Habenicht, A., ed. (1990) Growth Factors, Differentiation Factors, and Cytokines, Springer-Verlag, New York, NY.)

Growth and differentiation factors play critical roles in neoplastic transformation of cells in vitro and in tumor progression in vivo. Overexpression of the large polypeptide growth factors promotes the proliferation and transformation of cells in culture. Moreover, tumor cells and normal cells show differential expression of certain growth factors. For example, northern analysis shows the expression of human hepatoma-derived growth factor messenger RNA at various levels in hepatoma and other transformed cell lines and in normal tissues. (Nakamura, H. et al. (1994) J. Biol. Chem. 269:25143-25149.) Inappropriate expression of large polypeptide growth factors by tumor cells in vivo may contribute to vascularization and metastasis of tumors. Furthermore, some of the large polypeptide growth factors are both structurally and functionally related to oncoproteins, the cancer-correlated products of oncogenes. Certain FGF and PDGF family members are themselves homologous to oncoproteins, whereas receptors for some members of the EGF, NGF, and FGF families are encoded by proto-oncogenes. Growth factors also affect the transcriptional regulation of both proto-oncogenes and oncosuppressor genes. (Pimentel, supra.)

In addition, some of the large polypeptide growth factors play essential roles in embryonic development. In the fruit fly Drosophila melanogaster, for example, Twisted Gastrulation (TSG) is a secreted protein similar to human connective tissue growth factor, a PDGF-related protein induced by TGF- β activity. TSG protein is required for the specification of certain dorsal cell fates in the fruit fly embryo. This specification contributes to the establishment of the dorsoventral axis of the embryo. The dorsoventral axis, in turn, provides the framework for the entire adult body plan. Disruption of TSG

activity is lethal with the manifestation of developmental defects commencing at the time of gastrulation. (Mason, E. D. et al. (1994) *Genes Dev.* 8:1489-1501; and Zusman, S. B. and Weischaus, E. F. (1985) *Dev. Biol.* 111:359-371.)

The discovery of new human growth factor homologs and the polynucleotides
5 encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, treatment, and prevention of cell proliferative and developmental disorders.

SUMMARY OF THE INVENTION

10 The invention features substantially purified polypeptides, human growth factor homologs, referred to collectively as "HGFH" and individually as "HGFH-1," "HGFH-2," "HGFH-3," and "HGFH-4." In one aspect, the invention provides a substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1,
15 a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5, and a fragment of SEQ ID NO:7.

The invention further provides a substantially purified variant having at least 90% amino acid identity to the amino acid sequences of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7, or to a fragment of any of these sequences. The invention also
20 provides an isolated and purified polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5, and a fragment of SEQ ID NO:7. The invention also includes an isolated and purified polynucleotide variant having at least 90% polynucleotide
25 sequence identity to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5, and a fragment of SEQ ID NO:7.

Additionally, the invention provides an isolated and purified polynucleotide which
30 hybridizes under stringent conditions to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of

SEQ ID NO:3, a fragment of SEQ ID NO:5, and a fragment of SEQ ID NO:7, as well as an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7,
5 a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5, and a fragment of SEQ ID NO:7.

The invention also provides an isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, a fragment of SEQ ID NO:2, a fragment of SEQ ID
10 NO:4, a fragment of SEQ ID NO:6, and a fragment of SEQ ID NO:8. The invention further provides an isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide sequence comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, a fragment of SEQ ID NO:2, a fragment of SEQ ID
15 NO:4, a fragment of SEQ ID NO:6, and a fragment of SEQ ID NO:8, as well as an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, a fragment of SEQ ID NO:2, a fragment of SEQ ID NO:4, a fragment of SEQ ID NO:6, and a fragment of SEQ
20 ID NO:8. The invention also provides a polynucleotide fragment useful for designing oligonucleotides or to be used as a hybridization probe comprising a polynucleotide sequence selected from the group consisting of nucleotides 535-570 of SEQ ID NO:2, nucleotides 124-156 of SEQ ID NO:4, nucleotides 30-62 of SEQ ID NO:6, and nucleotides 147-182 of SEQ ID NO:8.

25 The invention further provides an expression vector containing at least a fragment of the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5, and a fragment of SEQ ID NO:7. In another aspect, the expression vector is
30 contained within a host cell.

The invention also provides a method for producing a polypeptide comprising the amino acid sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a

fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5, or a fragment of SEQ ID NO:7, the method comprising the steps of: (a) culturing the host cell containing an expression vector containing at least a fragment of a polynucleotide encoding the polypeptide under conditions suitable for the expression of the polypeptide;
5 and (b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5, or a fragment of SEQ ID NO:7 in conjunction with
10 a suitable pharmaceutical carrier.

The invention further includes a purified antibody which binds to a polypeptide comprising the amino acid sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5, or a fragment of SEQ ID NO:7, as well as a purified agonist and a purified
15 antagonist to the polypeptide. The invention also provides a method for treating or preventing a cell proliferative disorder, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of HGFH. The invention also provides a method for treating or preventing a developmental disorder, the method comprising administering to a subject in need of such treatment an effective amount of a
20 pharmaceutical composition comprising substantially purified HGFH.

The invention also provides a method for detecting a polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5, or a fragment of SEQ ID NO:7 in a biological sample
25 containing nucleic acids, the method comprising the steps of: (a) hybridizing the complement of the polynucleotide sequence encoding the polypeptide comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5, or a fragment of SEQ ID NO:7 to at least one of the nucleic acids of the biological sample, thereby forming a hybridization
30 complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide encoding the polypeptide in the biological sample. In one aspect, the nucleic acids of the biological

sample are amplified by the polymerase chain reaction prior to the hybridizing step.

BRIEF DESCRIPTION OF THE FIGURES

Figures 1A, 1B, and 1C show the amino acid sequence alignment between HGFH-
1 (862403; SEQ ID NO:1) and mouse growth and differentiation factor 9 (GI 567206;
5 SEQ ID NO:9).

Figures 2A and 2B show the amino acid sequence alignment between HGFH-2
(2676869; SEQ ID NO:3) and human hepatoma-derived growth factor (GI 598956; SEQ
ID NO:10).

10 Figures 3A, 3B, 3C, and 3D show the amino acid sequence alignment between
HGFH-3 (1568019; SEQ ID NO:5) and yeast Yhr074wp (GI 500832; SEQ ID NO:11).

Figures 4A and 4B show the amino acid sequence alignment between HGFH-4
(3577857; SEQ ID NO:7) and fruit fly Twisted Gastrulation (GI 529900; SEQ ID NO:12).

The alignments were produced using the multisequence alignment program of
15 DNASTAR™ software (DNASTAR Inc, Madison WI).

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is
understood that this invention is not limited to the particular methodology, protocols, cell
20 lines, vectors, and reagents described, as these may vary. It is also to be understood that
the terminology used herein is for the purpose of describing particular embodiments only,
and is not intended to limit the scope of the present invention which will be limited only
by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms
25 "a," "an," and "the" include plural reference unless the context clearly dictates otherwise.
Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a
reference to "an antibody" is a reference to one or more antibodies and equivalents thereof
known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the
30 same meanings as commonly understood by one of ordinary skill in the art to which this
invention belongs. Although any methods and materials similar or equivalent to those
described herein can be used in the practice or testing of the present invention, the

preferred methods, devices, and materials are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, vectors, and methodologies which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the
5 invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"HGFH," as used herein, refers to the amino acid sequences of substantially purified HGFH obtained from any species, particularly a mammalian species, including
10 bovine, ovine, porcine, murine, equine, and preferably the human species, from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist," as used herein, refers to a molecule which, when bound to HGFH, increases or prolongs the duration of the effect of HGFH. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate
15 the effect of HGFH.

An "allele" or an "allelic sequence," as these terms are used herein, is an alternative form of the gene encoding HGFH. Alleles may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. Any given natural or recombinant
20 gene may have none, one, or many allelic forms. Common mutational changes which give rise to alleles are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding HGFH, as described herein, include
25 those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide the same HGFH or a polypeptide with at least one functional characteristic of HGFH. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding HGFH, and improper or unexpected hybridization to alleles, with
30 a locus other than the normal chromosomal locus for the polynucleotide sequence encoding HGFH. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result

in a functionally equivalent HGFH. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of HGFH is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine.

The terms "amino acid" or "amino acid sequence," as used herein, refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. In this context, "fragments", "immunogenic fragments", or "antigenic fragments" refer to fragments of HGFH which are preferably about 5 to about 15 amino acids in length and which retain some biological activity or immunological activity of HGFH. Where "amino acid sequence" is recited herein to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification," as used herein, relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art. (See, e.g., Dieffenbach, C.W. and G.S. Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, pp.1-5.)

The term "antagonist," as it is used herein, refers to a molecule which, when bound to HGFH, decreases the amount or the duration of the effect of the biological or immunological activity of HGFH. Antagonists may include proteins, nucleic acids, carbohydrates, antibodies, or any other molecules which decrease the effect of HGFH.

As used herein, the term "antibody" refers to intact molecules as well as to fragments thereof, such as Fa, F(ab')₂, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies that bind HGFH polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized

chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

5 The term "antigenic determinant," as used herein, refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (given regions or three-dimensional structures on the protein). An antigenic
10 determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

 The term "antisense," as used herein, refers to any composition containing a nucleic acid sequence which is complementary to a specific nucleic acid sequence. The term "antisense strand" is used in reference to a nucleic acid strand that is complementary
15 to the "sense" strand. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" can refer to the antisense strand, and the designation "positive" can refer to the sense strand.

20 As used herein, the term "biologically active," refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic HGFH, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

25 The terms "complementary" or "complementarity," as used herein, refer to the natural binding of polynucleotides under permissive salt and temperature conditions by base pairing. For example, the sequence "A-G-T" binds to the complementary sequence "T-C-A." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total
30 complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular

importance in amplification reactions, which depend upon binding between nucleic acids strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" or a "composition comprising a given amino acid sequence," as these terms are used herein, refer broadly to
5 any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation, an aqueous solution, or a sterile composition. Compositions comprising polynucleotide sequences encoding HGFH or fragments of HGFH may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate.
10 In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

The phrase "consensus sequence," as used herein, refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using XL-PCR™ (Perkin
15 Elmer, Norwalk, CT) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of more than one Incyte Clone using a computer program for fragment assembly, such as the GELVIEW™ Fragment Assembly system (GCG, Madison, WI). Some sequences have been both extended and assembled to produce the consensus sequence .

20 As used herein, the term "correlates with expression of a polynucleotide" indicates that the detection of the presence of nucleic acids, the same or related to a nucleic acid sequence encoding HGFH, by northern analysis is indicative of the presence of nucleic acids encoding HGFH in a sample, and thereby correlates with expression of the transcript from the polynucleotide encoding HGFH.

25 A "deletion," as the term is used herein, refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative," as used herein, refers to the chemical modification of HGFH, of a polynucleotide sequence encoding HGFH, or of a polynucleotide sequence
30 complementary to a polynucleotide sequence encoding HGFH. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at

least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

- 5 The term "homology," as used herein, refers to a degree of complementarity. There may be partial homology or complete homology. The word "identity" may substitute for the word "homology." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially-homologous." The inhibition of hybridization of the completely
- 10 complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially homologous sequence or hybridization probe will compete for and inhibit the binding of a completely homologous sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of
- 15 reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% homology or identity). In the absence of non-specific binding, the
- 20 substantially homologous sequence or probe will not hybridize to the second non-complementary target sequence.

- The phrases "percent identity" or "% identity" refer to the percentage of sequence similarity found in a comparison of two or more amino acid or nucleic acid sequences. Percent identity can be determined electronically, e.g., by using the MegAlign program
- 25 (Lasergene software package, DNASTAR, Inc., Madison WI). The MegAlign program can create alignments between two or more sequences according to different methods, e.g., the Clustal Method. (Higgins, D.G. and P. M. Sharp (1988) Gene 73:237-244.) The Clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. The percentage similarity
- 30 between two amino acid sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between

sequence A and sequence B, times one hundred. Gaps of low or of no homology between the two amino acid sequences are not included in determining percentage similarity.

Percent identity between nucleic acid sequences can also be calculated by the Clustal Method, or by other methods known in the art, such as the Jotun Hein Method. (See, e.g.,
5 Hein, J. (1990) Methods in Enzymology 183:626-645.) Identity between sequences can also be determined by other methods known in the art, e.g., by varying hybridization conditions.

“Human artificial chromosomes” (HACs), as described herein, are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size,
10 and which contain all of the elements required for stable mitotic chromosome segregation and maintenance. (See, e.g., Harrington, J.J. et al. (1997) Nat Genet. 15:345-355.)

The term “humanized antibody,” as used herein, refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original
15 binding ability.

“Hybridization,” as the term is used herein, refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

As used herein, the term “hybridization complex” as used herein, refers to a complex formed between two nucleic acid sequences by virtue of the formation of
20 hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C₀t or R₀t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

25 The words “insertion” or “addition,” as used herein, refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, to the sequence found in the naturally occurring molecule.

“Immune response” can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be
30 characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

The term “microarray,” as used herein, refers to an array of distinct polynucleotides

or oligonucleotides arrayed on a substrate, such as paper, nylon or any other type of membrane, filter, chip, glass slide, or any other suitable solid support.

The term "modulate," as it appears herein, refers to a change in the activity of HGFH. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of HGFH.

The phrases "nucleic acid" or "nucleic acid sequence," as used herein, refer to an oligonucleotide, nucleotide, polynucleotide, or any fragment thereof, to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material. In this context, "fragments" refers to those nucleic acid sequences which are greater than about 60 nucleotides in length, and most preferably are at least about 100 nucleotides, at least about 1000 nucleotides, or at least about 10,000 nucleotides in length.

The terms "operably associated" or "operably linked," as used herein, refer to functionally related nucleic acid sequences. A promoter is operably associated or operably linked with a coding sequence if the promoter controls the transcription of the encoded polypeptide. While operably associated or operably linked nucleic acid sequences can be contiguous and in reading frame, certain genetic elements, e.g., repressor genes, are not contiguously linked to the encoded polypeptide but still bind to operator sequences that control expression of the polypeptide.

The term "oligonucleotide," as used herein, refers to a nucleic acid sequence of at least about 6 nucleotides to 60 nucleotides, preferably about 15 to 30 nucleotides, and most preferably about 20 to 25 nucleotides, which can be used in PCR amplification or in a hybridization assay or microarray. As used herein, the term "oligonucleotide" is substantially equivalent to the terms "amplimers," "primers," "oligomers," and "probes," as these terms are commonly defined in the art.

"Peptide nucleic acid" (PNA), as used herein, refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA and RNA and stop transcript elongation, and may be pegylated to

extend their lifespan in the cell. (See, e.g., Nielsen, P.E. et al. (1993) Anticancer Drug Des. 8:53-63.)

The term "sample," as used herein, is used in its broadest sense. A biological sample suspected of containing nucleic acids encoding HGFH, or fragments thereof, or HGFH itself may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a solid support; a tissue; a tissue print; etc.

As used herein, the terms "specific binding" or "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, or an antagonist. The interaction is dependent upon the presence of a particular structure of the protein recognized by the binding molecule (i.e., the antigenic determinant or epitope). For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

As used herein, the term "stringent conditions" refers to conditions which permit hybridization between polynucleotide sequences and the claimed polynucleotide sequences. Suitably stringent conditions can be defined by, for example, the concentrations of salt or formamide in the prehybridization and hybridization solutions, or by the hybridization temperature, and are well known in the art. In particular, stringency can be increased by reducing the concentration of salt, increasing the concentration of formamide, or raising the hybridization temperature.

For example, hybridization under high stringency conditions could occur in about 50% formamide at about 37°C to 42°C. Hybridization could occur under reduced stringency conditions in about 35% to 25% formamide at about 30°C to 35°C. In particular, hybridization could occur under high stringency conditions at 42°C in 50% formamide, 5X SSPE, 0.3% SDS, and 200 $\mu\text{g/ml}$ sheared and denatured salmon sperm DNA. Hybridization could occur under reduced stringency conditions as described above, but in 35% formamide at a reduced temperature of 35°C. The temperature range corresponding to a particular level of stringency can be further narrowed by calculating the purine to pyrimidine ratio of the nucleic acid of interest and adjusting the temperature accordingly. Variations on the above ranges and conditions are well known in the art.

The term "substantially purified," as used herein, refers to nucleic acid or amino

acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which they are naturally associated.

A "substitution," as used herein, refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Transformation," as defined herein, describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, and refers to cells which transiently express the inserted DNA or RNA for limited periods of time.

A "variant" of HGFH, as used herein, refers to an amino acid sequence that is altered by one or more amino acids. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (e.g., replacement of leucine with isoleucine). More rarely, a variant may have "nonconservative" changes (e.g., replacement of glycine with tryptophan). Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, DNASTAR software.

THE INVENTION

The invention is based on the discovery of new human growth factor homologs (HGFH), the polynucleotides encoding HGFH, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative and developmental disorders.

Nucleic acids encoding the HGFH-1 of the present invention were first identified in Incyte Clone 862403 from the brain tumor cDNA library (BRAITUT03) using a

computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:2, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 862403 and 860725 (BRAITUT03) and shotgun sequences SAAB00148, SAAC00081, SAAB00364, SAAB00198, and SZZZ00645.

- 5 In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:1. HGFH-1 is 446 amino acids in length and has five potential N-glycosylation sites at N₁₀₆, N₂₂₈, N₂₄₇, N₂₆₀, and N₃₃₀; one potential cAMP- and cGMP-dependent protein kinase phosphorylation site at S₂₈₀; three potential casein kinase II phosphorylation sites at S₂₈, T₁₄₈, and S₁₉₃; six potential protein kinase C
- 10 phosphorylation sites at S₇₆, S₁₀₅, T₁₁₃, S₁₇₈, S₂₁₈, and T₄₄₄; a potential signal peptide sequence from M₁ to P₂₀; and a TGF- β family signature from I₃₆₃ to C₃₇₈. The region of HGFH-1 from C₃₄₅ to R₄₄₆ also shows similarity to a TGF- β protein domain. As shown in Figures 1A, 1B, and 1C, HGFH-1 has chemical and structural homology with mouse growth/differentiation factor 9 (GDF-9), a member of the TGF- β family (GI 567206; SEQ
- 15 ID NO:9). In particular, HGFH-1 and mouse GDF-9 share 66% identity overall and 95% identity in their C-terminal regions from A₃₂₇ to R₄₄₆ of HGFH-1. Within this region, cysteines important for intramolecular disulfide bond formation in GDF-9 are conserved in HGFH-1 at C₃₄₅, C₃₇₄, C₃₇₈, C₄₁₁, C₄₄₃, and C₄₄₅. S₄₀₅ of GDF-9, which distinguishes GDF-9 from other TGF- β -like proteins that contain a cysteine at this position, is conserved in
- 20 HGFH-1 at S₄₁₀. In addition, the potential N-glycosylation sites at N₂₂₈, N₂₆₀, and N₃₃₀; the potential phosphorylation sites at S₂₈, S₇₆, T₁₁₃, T₁₄₈, S₁₇₈, and T₄₄₄; and the potential signal peptide sequence of HGFH-1 are conserved in GDF-9. The fragment of SEQ ID NO:2 from about nucleotide 535 to nucleotide 570 is useful for designing oligonucleotides or to be used directly as a hybridization probe. Northern analysis shows the expression of this
- 25 sequence in a cDNA library derived from tumorous brain tissue.

- Nucleic acids encoding the HGFH-2 of the present invention were first identified in Incyte Clone 2676869 from the kidney cDNA library (KIDNNOT19) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:4, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte
- 30 Clones 2676869 (KIDNNOT19), 488482 (HNT2AGT01), 2571655 (HIPOAZT01), 2150727 (BRAINOT09), 899636 (BRSTTUT03), 2311937 (NGANNOT01), 2613941 (THYRNOT09), 776810 (COLNNOT05), 968484 (BRSTNOT05), 1292885

(PGANNOT03), and 2288692 (BRAINON01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:3. HGFH-2 is 203 amino acids in length and has three potential N-glycosylation sites at N₁₇₆, N₁₈₇, and N₁₉₁; two potential cAMP- and cGMP-dependent protein kinase phosphorylation sites at S₁₄₉ and S₁₅₆; eight potential casein kinase II phosphorylation sites at S₁₀₆, S₁₀₈, T₁₁₀, S₁₂₁, S₁₂₂, S₁₆₂, S₁₇₈, and T₁₉₂; and six potential protein kinase C phosphorylation sites at S₇₅, S₁₄₄, T₁₅₁, S₁₅₂, S₁₅₅, and S₁₅₉. As shown in Figures 2A and 2B, HGFH-2 has chemical and structural homology with human hepatoma-derived growth factor (HDGF) (GI 598956; SEQ ID NO:10). In particular, HGFH-2 and HDGF share 49% identity. In addition, the potential phosphorylation sites at S₁₀₆, S₁₂₂, S₁₅₂, and S₁₇₈ of HGFH-2 are conserved in HDGF. The fragment of SEQ ID NO:4 from about nucleotide 124 to nucleotide 156 is useful for designing oligonucleotides or to be used directly as a hybridization probe. Northern analysis shows the expression of this sequence in various libraries, at least 44% of which are associated with immortalized or cancerous tissue. Of particular note is the expression of HGFH-2 in neural tissue.

Nucleic acids encoding the HGFH-3 of the present invention were first identified in Incyte Clone 1568019 from the uterus cDNA library (UTRSNOT05) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:6, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 1568019 (UTRSNOT05), 2504437 (CONUTUT01), 136194 (SYNORAB01), 1849286 (LUNGFET03) and shotgun sequences SAEA02891, SAEA03437, and SAFC02217.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:5. HGFH-3 is 706 amino acids in length and has four potential N-glycosylation sites at N₁₉₈, N₂₅₁, N₄₂₂, and N₆₇₇; one potential cAMP- and cGMP-dependent protein kinase phosphorylation site at T₆; sixteen potential casein kinase II phosphorylation sites at T₁₅₁, T₂₁₁, S₂₆₂, S₃₀₈, S₃₂₇, S₃₈₂, T₄₀₁, S₄₂₃, S₄₇₄, T₅₂₇, S₅₃₂, S₅₄₃, S₅₈₄, T₅₈₆, T₅₉₄, and S₆₆₃; and six potential protein kinase C phosphorylation sites at T₂₂₀, S₂₈₀, S₂₈₆, T₄₆₂, S₄₇₄, and T₆₂₈. As shown in Figures 3A, 3B, 3C, and 3D, HGFH-3 has chemical and structural homology with yeast Yhr074wp (GI 500832; SEQ ID NO:11), an open reading frame with similarity to prokaryotic sporulation outgrowth factor B. In particular, HGFH-3 and Yhr074wp share 57% identity. In addition, the potential N-glycosylation

sites at N₁₉₈, N₂₅₁, and N₄₂₂, and the potential phosphorylation sites at T₂₂₀, S₂₆₂, S₂₈₀, S₃₂₇, S₄₂₃, T₄₆₂, S₄₇₄, T₅₂₇, S₅₃₂, S₅₄₃, T₅₉₄, T₆₂₈, and S₆₆₃ of HGFH-3 are conserved in Yhr074wp. The fragment of SEQ ID NO:6 from about nucleotide 30 to nucleotide 62 is useful for designing oligonucleotides or to be used directly as a hybridization probe. Northern
 5 analysis shows the expression of this sequence in various cDNA libraries, at least 39% of which are associated with cancerous or immortalized tissue and at least 31% with the immune response. Of particular note is the expression of HGFH-3 in reproductive and gastrointestinal tissues.

Nucleic acids encoding the HGFH-4 of the present invention were first identified
 10 in Incyte Clone 3577857 from the bronchial tissue cDNA library BRONNOT01 using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:8, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 3577857 (BRONNOT01), 2799484 (NPOLNOT01), 1685583 (PROSNOT15), 1210438 (BRSTNOT02), 3601074 (DRGTNOT01), 3279304
 15 (STOMFET02), and shotgun sequences SAFC00098, SAYA00353, and SAFC00552.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:7. HGFH-4 is 223 amino acids in length and has three potential N-glycosylation sites at N₅₂, N₈₁, and N₁₄₇; six potential casein kinase II phosphorylation sites at S₉₁, T₉₂, T₁₀₈, S₁₂₈, S₁₃₅, and S₁₆₁; three potential protein kinase C
 20 phosphorylation sites at T₈₈, S₁₆₁, and T₂₁₅; and a potential signal peptide sequence from M₁ to S₂₅. As shown in Figures 4A and 4B, HGFH-4 has chemical and structural homology with fruit fly Twisted Gastrulation (TSG) (GI 529900; SEQ ID NO:12). In particular, HGFH-4 and TSG share 29% amino acid identity. In addition, all 24 cysteine residues in HGFH-4 are conserved in TSG. The potential signal peptide sequence, acidic
 25 isoelectric point, and high cysteine content (~10%) of TSG are conserved in HGFH-4. The potential phosphorylation sites at T₈₈ and T₁₀₈ of HGFH-4 are conserved in TSG. The fragment of SEQ ID NO:8 from about nucleotide 147 to nucleotide 182 is useful for designing oligonucleotides or to be used directly as a hybridization probe. Northern analysis shows the expression of HGFH-4 in various libraries, at least 45% of which are
 30 associated with cancerous or immortalized tissue and at least 24% with immune response. Of particular note is the expression of HGFH-4 in reproductive and neural tissues.

The invention also encompasses HGFH variants. A preferred HGFH variant is one

which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the HGFH amino acid sequence, and which contains at least one functional or structural characteristic of HGFH.

The invention also encompasses polynucleotides which encode HGFH. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising the sequence of SEQ ID NO:2, which encodes an HGFH-1. In another embodiment, the invention encompasses a polynucleotide sequence comprising the sequence of SEQ ID NO:4, which encodes an HGFH-2. In a further embodiment, the invention encompasses a polynucleotide sequence comprising the sequence of SEQ ID NO:6, which encodes an HGFH-3. In still another embodiment, the invention encompasses a polynucleotide sequence comprising the sequence of SEQ ID NO:8, which encodes an HGFH-4.

The invention also encompasses a variant of a polynucleotide sequence encoding HGFH. In particular, such a variant polynucleotide sequence will have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding HGFH. A particular aspect of the invention encompasses a variant of SEQ ID NO:2 which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to SEQ ID NO:2. The invention further encompasses a polynucleotide variant of SEQ ID NO:4 having at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to SEQ ID NO:4. The invention further encompasses a polynucleotide variant of SEQ ID NO:6 having at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to SEQ ID NO:6. The invention further encompasses a polynucleotide variant of SEQ ID NO:8 having at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to SEQ ID NO:8. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of HGFH.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding HGFH, some bearing minimal homology to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible

variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring HGFH, and all such variations are to be considered as being specifically disclosed.

5 Although nucleotide sequences which encode HGFH and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring HGFH under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding HGFH or its derivatives possessing a substantially different codon usage. Codons may be selected to increase the rate at which expression of
10 the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding HGFH and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced
15 from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode HGFH and HGFH derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents that are well known in the art.
20 Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding HGFH or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, a fragment of SEQ ID
25 NO:2, a fragment of SEQ ID NO:4, a fragment of SEQ ID NO:6 or a fragment of SEQ ID NO:8, under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; and Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.)

Methods for DNA sequencing are well known and generally available in the art
30 and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase® (US Biochemical Corp., Cleveland, OH), Taq polymerase (Perkin Elmer), thermostable T7

polymerase (Amersham, Chicago, IL), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE Amplification System (GIBCO/BRL, Gaithersburg, MD). Preferably, the process is automated with machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno, NV), Peltier Thermal Cycler (PTC200; MJ Research, Watertown, MA) and the ABI Catalyst and 373 and 377 DNA Sequencers (Perkin Elmer).

The nucleic acid sequences encoding HGFH may be extended utilizing a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal primers to retrieve unknown sequence adjacent to a known locus. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) In particular, genomic DNA is first amplified in the presence of a primer complementary to a linker sequence within the vector and a primer specific to the region predicted to encode the gene. The amplified sequences are then subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR may also be used to amplify or extend sequences using divergent primers based on a known region. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) The primers may be designed using commercially available software such as OLIGO 4.06 Primer Analysis software (National Biosciences Inc., Plymouth, MN) or another appropriate program to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to 72°C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Another method which may be used is capture PCR, which involves PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to place an engineered double-stranded sequence into an unknown fragment of the DNA molecule before performing PCR. Other methods which may be used to

retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060.) Additionally, one may use PCR, nested primers, and PromoterFinder™ libraries to walk genomic DNA (Clontech, Palo Alto, CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

5 When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable in that they will include more sequences which contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension
10 of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different fluorescent dyes (one for each nucleotide) which are laser
15 activated, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., Genotyper™ and Sequence Navigator™, Perkin Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for the sequencing of small
20 pieces of DNA which might be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode HGFH may be used in recombinant DNA molecules to direct expression of HGFH, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which
25 encode substantially the same or a functionally equivalent amino acid sequence may be produced, and these sequences may be used to clone and express HGFH.

As will be understood by those of skill in the art, it may be advantageous to produce HGFH-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can
30 be selected to increase the rate of protein expression or to produce an RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter HGFH-encoding sequences for a variety of reasons including, but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. DNA shuffling by random

5 fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, introduce mutations, and so forth.

In another embodiment of the invention, natural, modified, or recombinant nucleic
10 acid sequences encoding HGFH may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of HGFH activity, it may be useful to encode a chimeric HGFH protein that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between the HGFH encoding sequence and the heterologous protein
15 sequence, so that HGFH may be cleaved and purified away from the heterologous moiety.

In another embodiment, sequences encoding HGFH may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucl. Acids Res. Symp. Ser. 215-223, and Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser. 225-232.) Alternatively, the protein itself may be produced using chemical
20 methods to synthesize the amino acid sequence of HGFH, or a fragment thereof. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A Peptide Synthesizer (Perkin Elmer).

The newly synthesized peptide may be substantially purified by preparative high
25 performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1983) Proteins, Structures and Molecular Properties, WH Freeman and Co., New York, NY.) Additionally, the amino acid sequence of HGFH, or any part thereof, may be altered
30 during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

In order to express a biologically active HGFH, the nucleotide sequences encoding

HGFH or derivatives thereof may be inserted into appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence.

Methods which are well known to those skilled in the art may be used to construct
5 expression vectors containing sequences encoding HGFH and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning. A Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, ch. 4, 8, and 16-17; and Ausubel, F.M. et al. (1995, and periodic
10 supplements) Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding HGFH. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression
15 vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus (CaMV) or tobacco mosaic virus (TMV)) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems.

20 The invention is not limited by the host cell employed.

The "control elements" or "regulatory sequences" are those non-translated regions, e.g., enhancers, promoters, and 5' and 3' untranslated regions, of the vector and polynucleotide sequences encoding HGFH which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and
25 specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters, e.g., hybrid lacZ promoter of the Bluescript® phagemid (Stratagene, La Jolla, CA) or pSport1™ plasmid (GIBCO/BRL), may be used. The baculovirus polyhedrin promoter may be used
30 in insect cells. Promoters or enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO, and storage protein genes) or from plant viruses (e.g., viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems,

promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding HGFH, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

5 In bacterial systems, a number of expression vectors may be selected depending upon the use intended for HGFH. For example, when large quantities of HGFH are needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, multifunctional E. coli cloning and expression vectors such as Bluescript® (Stratagene),
10 in which the sequence encoding HGFH may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of β -galactosidase so that a hybrid protein is produced, and pIN vectors. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) pGEX vectors (Pharmacia Biotech, Uppsala, Sweden) may also be used to express foreign polypeptides as fusion proteins
15 with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

20 In the yeast Saccharomyces cerevisiae, a number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH, may be used. (See, e.g., Ausubel, supra; and Grant et al. (1987) Methods Enzymol. 153:516-544.)

In cases where plant expression vectors are used, the expression of sequences encoding HGFH may be driven by any of a number of promoters. For example, viral
25 promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV. (Takamatsu, N. (1987) EMBO J. 6:307-311.) Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991)
30 Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews. (See, e.g., Hobbs, S. or Murry, L.E.

in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, NY; pp. 191-196.)

An insect system may also be used to express HGFH. For example, in one such system, Autographa californica nuclear polyhedrosis virus (AcNPV) is used as a vector to
5 express foreign genes in Spodoptera frugiperda cells or in Trichoplusia larvae. The sequences encoding HGFH may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of sequences encoding HGFH will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be
10 used to infect, for example, S. frugiperda cells or Trichoplusia larvae in which HGFH may be expressed. (See, e.g., Engelhard, E.K. et al. (1994) Proc. Nat. Acad. Sci. 91:3224-3227.)

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding
15 HGFH may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing HGFH in infected host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. 81:3655-3659.) In addition, transcription enhancers, such as the Rous
20 sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes,
25 polycationic amino polymers, or vesicles) for therapeutic purposes.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding HGFH. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding HGFH and its initiation codon and upstream sequences are inserted into the appropriate expression vector, no additional
30 transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation

codon should be in the correct reading frame to ensure translation of the entire insert.

Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular cell system used. (See, e.g., Scharf, D. et al.

5 (1994) Results Probl. Cell Differ. 20:125-162.)

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing
10 which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding, and/or function. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC, Bethesda, MD) and may be chosen to ensure the correct modification
15 and processing of the foreign protein.

For long term, high yield production of recombinant proteins, stable expression is preferred. For example, cell lines capable of stably expressing HGFH can be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector.
20 Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to
25 the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase genes and adenine phosphoribosyltransferase genes, which can be employed in *tk* or *apv* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; and Lowy, I. et al.
30 (1980) Cell 22:817-823) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *npt* confers resistance to the aminoglycosides neomycin and G-418; and *als* or *pat* confer

resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-3570; Colbere-Garapin, F. et al (1981) J. Mol. Biol. 150:1-14; and Murry, supra.) Additional selectable genes have been described, e.g., *trpB*, which allows cells to utilize indole in place of tryptophan, or *hisD*, which allows cells to utilize histinol in place of histidine. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-8051.) Recently, the use of visible markers has gained popularity with such markers as anthocyanins, β glucuronidase and its substrate GUS, luciferase and its substrate luciferin. Green fluorescent proteins (GFP) (Clontech, Palo Alto, CA) are also used (See, e.g., Chalfie, M. et al. (1994) Science 263:802-805.) These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. et al. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding HGFH is inserted within a marker gene sequence, transformed cells containing sequences encoding HGFH can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding HGFH under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells which contain the nucleic acid sequence encoding HGFH and express HGFH may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

The presence of polynucleotide sequences encoding HGFH can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes or fragments or fragments of polynucleotides encoding HGFH. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the sequences encoding HGFH to detect transformants containing DNA or RNA encoding HGFH.

A variety of protocols for detecting and measuring the expression of HGFH, using

either polyclonal or monoclonal antibodies specific for the protein, are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two
5 non-interfering epitopes on HGFH is preferred, but a competitive binding assay may be employed. These and other assays are well described in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN, Section IV; and Maddox, D.E. et al. (1983) *J. Exp. Med.* 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in
10 the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding HGFH include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding HGFH, or any fragments thereof, may be cloned into a vector for the production of an
15 mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Pharmacia & Upjohn (Kalamazoo, MI), Promega (Madison, WI), and U.S. Biochemical Corp. (Cleveland, OH).
20 Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding HGFH may be cultured under conditions suitable for the expression and recovery of the protein from cell culture.
25 The protein produced by a transformed cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode HGFH may be designed to contain signal sequences which direct secretion of HGFH through a prokaryotic or eukaryotic cell membrane. Other constructions may be used to join
30 sequences encoding HGFH to nucleotide sequences encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan

modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, WA). The inclusion of cleavable linker sequences, such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, CA), between the purification domain and the HGFH encoding sequence may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing HGFH and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on immobilized metal ion affinity chromatography. (IMAC) (See, e.g., Porath, J. et al. (1992) Prot. Exp. Purif. 3: 263-281.) The enterokinase cleavage site provides a means for purifying HGFH from the fusion protein. (See, e.g., Kroll, D.J. et al. (1993) DNA Cell Biol. 12:441-453.)

Fragments of HGFH may be produced not only by recombinant production, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, T.E. (1984) Protein: Structures and Molecular Properties, pp. 55-60, W.H. Freeman and Co., New York, NY.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Various fragments of HGFH may be synthesized separately and then combined to produce the full length molecule.

THERAPEUTICS

Chemical and structural homology exists between HGFH-1 and GDF-9 from mouse (GI 567206). In addition, HGFH-1 is expressed in cancerous tissues. Therefore, HGFH-1 appears to play a role in cell proliferative and developmental disorders.

Chemical and structural homology exists between HGFH-2 and HDGF from human (GI 598956). In addition, HGFH-2 is expressed in cancerous tissues. Therefore, HGFH-2 appears to play a role in cell proliferative and developmental disorders.

Chemical and structural homology exists between HGFH-3 and Yhr074wp from yeast (GI 500832). In addition, HGFH-3 is expressed in cancerous tissues. Therefore, HGFH-3 appears to play a role in cell proliferative and developmental disorders.

Chemical and structural homology exists between HGFH-4 and TSG from fruit fly (GI 529900). In addition, HGFH-4 is expressed in cancerous tissues. Therefore, HGFH-4

appears to play a role in cell proliferative and developmental disorders.

Therefore, in one embodiment, an antagonist of HGFH may be administered to a subject to treat or prevent a cell proliferative disorder. Such a disorder may include, but is not limited to, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed
5 connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle,
10 ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. In one aspect, an antibody which specifically binds HGFH may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express HGFH.

In an additional embodiment, a vector expressing the complement of the
15 polynucleotide encoding HGFH may be administered to a subject to treat or prevent a cell proliferative disorder including, but not limited to, those described above.

In a further embodiment, HGFH or a fragment or derivative thereof may be administered to a subject to treat or prevent a developmental disorder. Such a disorder can include, but is not limited to, renal tubular acidosis, anemia, Cushing's syndrome,
20 achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome, Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spinal bifida, and
25 congenital glaucoma, cataract, or sensorineural hearing loss.

In another embodiment, a vector capable of expressing HGFH or a fragment or derivative thereof may be administered to a subject to treat or prevent a developmental disorder including, but not limited to, those described above.

In a still another embodiment, a pharmaceutical composition comprising a
30 substantially purified HGFH in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a developmental disorder including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of HGFH may be administered to a subject to treat or prevent a developmental disorder including, but not limited to, those listed above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, 5 complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described 10 above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of HGFH may be produced using methods which are generally known in the art. In particular, purified HGFH may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind HGFH. 15 Antibodies to HGFH may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

20 For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with HGFH or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and 25 surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to HGFH have an amino acid sequence consisting of at least about 5 amino 30 acids, and, more preferably, of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small,

naturally occurring molecule. Short stretches of HGFH amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to HGFH may be prepared using any technique which
5 provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci.* 80:2026-2030; and Cole, S.P. et al. (1984) *Mol.*
10 *Cell Biol.* 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) *Proc. Natl. Acad. Sci.* 81:6851-6855; Neuberger, M.S. et al.
15 (1984) *Nature* 312:604-608; and Takeda, S. et al. (1985) *Nature* 314:452-454.)

Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce HGFH-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries.
20 (See, e.g., Burton D.R. (1991) *Proc. Natl. Acad. Sci.* 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci.* 86: 3833-3837; and Winter, G. et al. (1991) *Nature* 349:293-299.)

25 Antibody fragments which contain specific binding sites for HGFH may also be generated. For example, such fragments include, but are not limited to, F(ab')₂ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')₂ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab
30 fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) *Science* 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the

desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between HGFH and its specific antibody. A two-site, monoclonal-based
5 immunoassay utilizing monoclonal antibodies reactive to two non-interfering HGFH epitopes is preferred, but a competitive binding assay may also be employed. (Maddox, supra.)

In another embodiment of the invention, the polynucleotides encoding HGFH, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect,
10 the complement of the polynucleotide encoding HGFH may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding HGFH. Thus, complementary molecules or fragments may be used to modulate HGFH activity, or to achieve regulation of gene function. Such technology is now well known in the art, and
15 sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding HGFH.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known
20 to those skilled in the art can be used to construct vectors which will express nucleic acid sequences complementary to the polynucleotides of the gene encoding HGFH. (See, e.g., Sambrook, supra; and Ausubel, supra.)

Genes encoding HGFH can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof,
25 encoding HGFH. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector
30 system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the

control, 5', or regulatory regions of the gene encoding HGFH. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding HGFH.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding HGFH. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell

lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life.

Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nature Biotechnology 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of HGFH, antibodies to HGFH, and mimetics, agonists, antagonists, or inhibitors of HGFH. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may

contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with

or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiologically buffered saline. Aqueous injection
5 suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino
10 polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the
15 art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

20 The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1
25 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of HGFH, such labeling would include amount, frequency, and method of
30 administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended

purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example HGFH or fragments thereof, antibodies of HGFH, and agonists, antagonists or inhibitors of HGFH, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED₅₀ (the dose therapeutically effective in 50% of the population) or LD₅₀ (the dose lethal to 50% of the population) statistics. The dose ratio of therapeutic to toxic effects is the therapeutic index, and it can be expressed as the ED₅₀/LD₅₀ ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED₅₀ with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μ g to 100,000 μ g, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different

formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

5 DIAGNOSTICS

In another embodiment, antibodies which specifically bind HGFH may be used for the diagnosis of disorders characterized by expression of HGFH, or in assays to monitor patients being treated with HGFH or agonists, antagonists, or inhibitors of HGFH.

Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for HGFH include methods which utilize the antibody and a label to detect HGFH in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

15 A variety of protocols for measuring HGFH, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of HGFH expression. Normal or standard values for HGFH expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to HGFH under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, preferably by photometric means. Quantities of HGFH expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

25 In another embodiment of the invention, the polynucleotides encoding HGFH may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of HGFH may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of HGFH, and to monitor regulation of HGFH levels during therapeutic intervention.

30 In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding HGFH or closely

related molecules may be used to identify nucleic acid sequences which encode HGFH.

The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low), will
5 determine whether the probe identifies only naturally occurring sequences encoding HGFH, alleles, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably contain at least 50% of the nucleotides from any of the HGFH encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and
10 may be derived from the sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, or from genomic sequences including promoters, enhancers, and introns of the HGFH gene.

Means for producing specific hybridization probes for DNAs encoding HGFH include the cloning of polynucleotide sequences encoding HGFH or HGFH derivatives
15 into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S , or by enzymatic labels, such as alkaline phosphatase
20 coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding HGFH may be used for the diagnosis of a disorder associated with expression of HGFH. Examples of such a disorder include, but are not limited to, a cell proliferative disorder, such as arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis,
25 paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid,
30 penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal

dysgenesis, WAGR syndrome, Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spinal bifida, and congenital glaucoma, cataract, or sensorineural hearing loss. The polynucleotide sequences encoding HGFH may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and ELISA assays; and in microarrays utilizing fluids or tissues from patients to detect altered HGFH expression. Such qualitative or quantitative methods are well known in the art.

10 In a particular aspect, the nucleotide sequences encoding HGFH may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding HGFH may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the
15 signal is quantitated and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding HGFH in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to
20 monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of HGFH, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding HGFH, under
25 conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard
30 values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of

expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of a relatively high amount of transcript in
5 biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

10 Additional diagnostic uses for oligonucleotides designed from the sequences encoding HGFH may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding HGFH, or a fragment of a polynucleotide complementary to the polynucleotide encoding HGFH, and will be employed under
15 optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of HGFH include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and
20 interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; and Duplaa, C. et al. (1993) Anal. Biochem. 229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

25 In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a
30 disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art.

(See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding HGFH may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, R.A. (ed.) Molecular Biology and Biotechnology, VCH Publishers New York, NY, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) site. Correlation between the location of the gene encoding HGFH on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., AT to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti,

R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, HGFH, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between HGFH and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with HGFH, or fragments thereof, and washed. Bound HGFH is then detected by methods well known in the art. Purified HGFH can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding HGFH specifically compete with a test compound for binding HGFH. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with HGFH.

In additional embodiments, the nucleotide sequences which encode HGFH may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention.

30

EXAMPLES

I. cDNA Library Construction

BRAITUT03

The BRAITUT03 cDNA library was constructed from brain tumor tissue removed from the left frontal lobe of a 17-year-old Caucasian female during excision of a cerebral meningeal lesion. The frozen tissue was homogenized and lysed in guanidinium

5 isothiocyanate solution using a Brinkmann Homogenizer Polytron-PT 3000 (Brinkmann Instruments, Westbury, NY). The lysate was extracted once with acid phenol, and RNA was isolated according to Stratagene's protocol (Stratagene, La Jolla, CA). The RNA was extracted twice with acid phenol, precipitated with sodium acetate and ethanol, resuspended in RNase-free water, and treated with DNase.

10 UTRSNOT05

The UTRSNOT05 cDNA library was constructed from nontumorous uterine tissue excised from a 45-year-old Caucasian female during a total abdominal hysterectomy and total colectomy. The frozen tissue was homogenized and lysed in guanidinium

15 isothiocyanate solution using a Brinkmann Homogenizer Polytron PT-3000 (Brinkmann Instruments). The lysate was centrifuged over a CsCl cushion to isolate RNA. The RNA was extracted with acid phenol, precipitated with sodium acetate and ethanol, resuspended in RNase-free water, and treated with DNase. The RNA was re-extracted and reprecipitated as described above. **KIDNNOT19 AND BRONNOT01**

The KIDNNOT19 cDNA library was constructed from nontumorous kidney tissue

20 removed from a 65-year-old Caucasian male during an exploratory laparotomy and nephroureterectomy. The BRONNOT01 cDNA library was constructed from bronchial tissue removed from a 15-year-old Caucasian male with a history of alcohol, tobacco, and marijuana use who died an accidental death. For each library, the frozen tissue was homogenized and lysed in TRIzol reagent (1 g tissue/10 ml TRIzol; Catalog #10296-028,

25 Gibco/BRL, Gaithersburg, MD), a monoplastic solution of phenol and guanidinium isothiocyanate, using a Brinkmann Homogenizer Polytron PT-3000 (Brinkmann Instruments). After a brief incubation on ice, chloroform was added (1:5 v/v), and the mixture was centrifuged. The upper aqueous layer was removed to a fresh tube, and the RNA was precipitated with isopropanol, resuspended in RNase-free water, and treated
30 with DNase. The RNA was reprecipitated using sodium acetate and ethanol. For BRONNOT01, RNA was extracted with acid phenol prior to this final precipitation step.

BRAITUT03, UTRSNOT05, KIDNNOT19, and BRONNOT01

From each of the four RNA preparations described above, Poly(A⁺) RNA was isolated using the Qiagen Oligotex kit (QIAGEN Inc, Chatsworth, CA). Poly(A⁺) RNA was used to construct each cDNA library according to the recommended protocols in the SuperScript Plasmid System (Catalog #18248-013, Gibco/BRL, Gaithersburg, MD). The

5 cDNAs were fractionated on a Sepharose CL4B column (Catalog #275105, Pharmacia, Piscataway, NJ). For BRAITUT03, those cDNAs exceeding 400 bp were ligated into the pSport I plasmid (Gibco/BRL). For UTRSNOT05, KIDNNOT19, AND BRONNOT01, those cDNAs exceeding 400 bp were ligated into the pINCY 1 plasmid (Incyte), a derivative of pSPORT I. All plasmids were subsequently transformed into DH5 α TM

10 competent cells (Catalog #18258-012, Gibco/BRL).

II. Isolation and Sequencing of cDNA Clones (All cDNA Libraries)

Plasmid DNA was released from the cells and purified using the REAL Prep 96 Plasmid Kit (Catalog #26173; QIAGEN Inc). The recommended protocol was employed

15 except for the following changes: 1) the bacteria were cultured in 1 ml of sterile Terrific Broth (Catalog #22711, Gibco/BRL) with carbenicillin at 25 mg/L and glycerol at 0.4%; 2) after inoculation, the cultures were incubated for 19 hours and then lysed with 0.3 ml of lysis buffer; and 3) following isopropanol precipitation, the plasmid DNA pellets were resuspended in 0.1 ml of distilled water. The plasmid DNA samples were stored at 4°C.

20 The cDNAs were sequenced by the method of Sanger et al. (1975, J. Mol. Biol. 94:441f), using a Hamilton Micro Lab 2200 (Hamilton, Reno, NV) in combination with Peltier Thermal Cyclers (PTC200 from MJ Research, Watertown, MA) and Applied Biosystems 377 DNA Sequencing Systems, and the reading frames were determined.

25 III. Homology Searching of cDNA Clones and Their Deduced Proteins

The nucleotide sequences and/or amino acid sequences of the Sequence Listing were used to query sequences in the GenBank, SwissProt, BLOCKS, and Pima II databases. These databases, which contain previously identified and annotated sequences, were searched for regions of homology using BLAST (Basic Local Alignment Search

30 Tool). (See, e.g., Altschul, S.F. (1993) J. Mol. Evol 36:290-300; and Altschul et al. (1990) J. Mol. Biol. 215:403-410.)

BLAST produced alignments of both nucleotide and amino acid sequences to

determine sequence similarity. Because of the local nature of the alignments, BLAST was especially useful in determining exact matches or in identifying homologs which may be of prokaryotic (bacterial) or eukaryotic (animal, fungal, or plant) origin. Other algorithms could have been used when dealing with primary sequence patterns and secondary structure gap penalties. (See, e.g., Smith, T. et al. (1992) Protein Engineering 5:35-51.) The sequences disclosed in this application have lengths of at least 49 nucleotides and have no more than 12% uncalled bases (where N is recorded rather than A, C, G, or T).

The BLAST approach searched for matches between a query sequence and a database sequence. BLAST evaluated the statistical significance of any matches found, and reported only those matches that satisfy the user-selected threshold of significance. In this application, threshold was set at 10^{-25} for nucleotides and 10^{-8} for peptides.

Incyte nucleotide sequences were searched against the GenBank databases for primate (pri), rodent (rod), and other mammalian sequences (mam), and deduced amino acid sequences from the same clones were then searched against GenBank functional protein databases, mammalian (mamp), vertebrate (vrtp), and eukaryote (eukp), for homology.

IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; and Ausubel, F.M. et al. *supra*, ch. 4 and 16.)

Analogous computer techniques applying BLAST are used to search for identical or related molecules in nucleotide databases such as GenBank or LIFESEQ™ database (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or homologous.

The basis of the search is the product score, which is defined as:

$$\frac{\% \text{ sequence identity} \times \% \text{ maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences

and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Homologous molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

5 The results of northern analysis are reported as a list of libraries in which the transcript encoding HGFH occurs. Abundance and percent abundance are also reported. Abundance directly reflects the number of times a particular transcript is represented in a cDNA library, and percent abundance is abundance divided by the total number of sequences examined in the cDNA library.

10

V. **Extension of HGFH Encoding Polynucleotides**

 The nucleic acid sequences of Incyte Clones 862403, 2676869, 1568019, and 3577857 were used to design oligonucleotide primers for extending partial nucleotide sequences to full length. For each nucleic acid sequence, one primer was synthesized to
15 initiate extension of an antisense polynucleotide, and the other was synthesized to initiate extension of a sense polynucleotide. Primers were used to facilitate the extension of the known sequence "outward" generating amplicons containing new unknown nucleotide sequence for the region of interest. The initial primers were designed from the cDNA using OLIGO 4.06 (National Biosciences, Plymouth, MN), or another appropriate
20 program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

 Selected human cDNA libraries (GIBCO/BRL) were used to extend the sequence.
25 If more than one extension is necessary or desired, additional sets of primers are designed to further extend the known region.

 High fidelity amplification was obtained by following the instructions for the XL-PCR kit (Perkin Elmer) and thoroughly mixing the enzyme and reaction mix. PCR was performed using the Peltier Thermal Cycler (PTC200; M.J. Research, Watertown, MA),
30 beginning with 40 pmol of each primer and the recommended concentrations of all other components of the kit, with the following parameters:

 Step 1 94° C for 1 min (initial denaturation)

	Step 2	65° C for 1 min
	Step 3	68° C for 6 min
	Step 4	94° C for 15 sec
	Step 5	65° C for 1 min
5	Step 6	68° C for 7 min
	Step 7	Repeat steps 4 through 6 for an additional 15 cycles
	Step 8	94° C for 15 sec
	Step 9	65° C for 1 min
	Step 10	68° C for 7:15 min
10	Step 11	Repeat steps 8 through 10 for an additional 12 cycles
	Step 12	72° C for 8 min
	Step 13	4° C (and holding)

A 5 μ l to 10 μ l aliquot of the reaction mixture was analyzed by electrophoresis on
 15 a low concentration (about 0.6% to 0.8%) agarose mini-gel to determine which reactions
 were successful in extending the sequence. Bands thought to contain the largest products
 were excised from the gel, purified using QIAQuick™ (QIAGEN Inc., Chatsworth, CA),
 and trimmed of overhangs using Klenow enzyme to facilitate religation and cloning.

After ethanol precipitation, the products were redissolved in 13 μ l of ligation
 20 buffer, 1 μ l T4-DNA ligase (15 units) and 1 μ l T4 polynucleotide kinase were added, and
 the mixture was incubated at room temperature for 2 to 3 hours, or overnight at 16° C.
 Competent *E. coli* cells (in 40 μ l of appropriate media) were transformed with 3 μ l of
 ligation mixture and cultured in 80 μ l of SOC medium. (See, e.g., Sambrook, *supra*,
 Appendix A, p. 2.) After incubation for one hour at 37° C, the *E. coli* mixture was plated
 25 on Luria Bertani (LB) agar (See, e.g., Sambrook, *supra*, Appendix A, p. 1) containing 2x
 Carb. The following day, several colonies were randomly picked from each plate and
 cultured in 150 μ l of liquid LB/2x Carb medium placed in an individual well of an
 appropriate commercially-available sterile 96-well microtiter plate. The following day, 5
 μ l of each overnight culture was transferred into a non-sterile 96-well plate and, after
 30 dilution 1:10 with water, 5 μ l from each sample was transferred into a PCR array.

For PCR amplification, 18 μ l of concentrated PCR reaction mix (3.3x) containing
 4 units of rTth DNA polymerase, a vector primer, and one or both of the gene specific
 primers used for the extension reaction were added to each well. Amplification was
 performed using the following conditions:

35	Step 1	94° C for 60 sec
	Step 2	94° C for 20 sec
	Step 3	55° C for 30 sec

- Step 4 72° C for 90 sec
Step 5 Repeat steps 2 through 4 for an additional 29 cycles
Step 6 72° C for 180 sec
Step 7 4° C (and holding)

5

Aliquots of the PCR reactions were run on agarose gels together with molecular weight markers. The sizes of the PCR products were compared to the original partial cDNAs, and appropriate clones were selected, ligated into plasmid, and sequenced.

In like manner, the nucleotide sequences of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, and SEQ ID NO:8 are used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for 5' extension, and an appropriate genomic library.

VI. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, and SEQ ID NO:8 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μ Ci of [γ -³²P] adenosine triphosphate (Amersham, Chicago, IL), and T4 polynucleotide kinase (DuPont NEN[®], Boston, MA). The labeled oligonucleotides are substantially purified using a Sephadex G-25 superfine resin column (Pharmacia & Upjohn, Kalamazoo, MI). An aliquot containing 10⁷ counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN, Boston, MA).

The DNA from each digest is fractionated on a 0.7 percent agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham, NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMAT AR[™] film (Kodak, Rochester, NY) is exposed to the blots to film for several hours, hybridization patterns are compared visually.

VII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, *supra*.) An array
5 analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using or thermal, UV, mechanical, or chemical bonding procedures, or a vacuum system. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns
10 of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

In another alternative, full-length cDNAs or Expressed Sequence Tags (ESTs) comprise the elements of the microarray. Full-length cDNAs or ESTs corresponding to
15 one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., U.V. cross-linking followed, by thermal and chemical and subsequent drying. (See, e.g., Schena, M. et al. (1995) *Science* 270:467-470; and Shalon, D. et al. (1996) *Genome Res.* 6:639-645.) Fluorescent
20 probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

Probe sequences for microarrays may be selected by screening a large number of clones from a variety of cDNA libraries in order to find sequences with conserved protein motifs common to genes coding for signal sequence containing polypeptides. In one
25 embodiment, sequences identified from cDNA libraries, are analyzed to identify those gene sequences with conserved protein motifs using an appropriate analysis program, e.g., the Block 2 Bioanalysis Program (Incyte, Palo Alto, CA). This motif analysis program, based on sequence information contained in the Swiss-Prot Database and PROSITE, is a method of determining the function of uncharacterized proteins translated from genomic or
30 cDNA sequences. (See, e.g., Bairoch, A. et al. (1997) *Nucleic Acids Res.* 25:217-221; and Attwood, T. K. et al. (1997) *J. Chem. Inf. Comput. Sci.* 37:417-424.) PROSITE may be used to identify functional or structural domains that cannot be detected using conserved

motifs due to extreme sequence divergence. The method is based on weight matrices. Motifs identified by this method are then calibrated against the SWISS-PROT database in order to obtain a measure of the chance distribution of the matches.

In another embodiment, Hidden Markov models (HMMs) may be used to find
5 shared motifs, specifically consensus sequences. (See, e.g., Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. 85:2444-2448; and Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197.) HMMs were initially developed to examine speech recognition patterns, but are now being used in a biological context to analyze protein and nucleic acid sequences as well as to model protein structure. (See, e.g., Krogh, A. et al.
10 (1994) J. Mol. Biol. 235:1501-1531; and Collin, M. et al. (1993) Protein Sci. 2:305-314.) HMMs have a formal probabilistic basis and use position-specific scores for amino acids or nucleotides. The algorithm continues to incorporate information from newly identified sequences to increase its motif analysis capabilities.

15 VIII. Complementary Polynucleotides

Sequences complementary to the HGFH-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring HGFH. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate
20 oligonucleotides are designed using Oligo 4.06 software and the coding sequence of HGFH. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the HGFH-encoding transcript.

25

IX. Expression of HGFH

Expression of HGFH is accomplished by subcloning the cDNA into an appropriate vector and transforming the vector into host cells. This vector contains an appropriate promoter, e.g., β -galactosidase upstream of the cloning site, operably
30 associated with the cDNA of interest. (See, e.g., Sambrook, *supra*, pp. 404-433; and Rosenberg, M. et al. (1983) Methods Enzymol. 101:123-138.)

Induction of an isolated, transformed bacterial strain with isopropyl beta-D-

thiogalactopyranoside (IPTG) using standard methods produces a fusion protein which consists of the first 8 residues of β -galactosidase, about 5 to 15 residues of linker, and the full length protein. The signal residues direct the secretion of HGFH into bacterial growth media which can be used directly in the following assay for activity.

5

X. Demonstration of HGFH Activity

The prototypical assay for growth factor activity measures the stimulation of DNA synthesis in Swiss mouse 3T3 cells. (McKay and Leigh, supra.) Initiation of DNA synthesis indicates the cells' entry into the mitotic cycle and their commitment to undergo later division. 3T3 cells are competent to respond to most growth factors, not only those that are mitogenic, but also those that are involved in embryonic induction. This competency is possible because the in vivo specificity demonstrated by some growth factors is not necessarily inherent but is determined by the responding tissue. Therefore, this assay is generally applicable to HGFH. In this assay, varying amounts of HGFH are added to quiescent 3T3 cultured cells in the presence of [3 H]thymidine, a radioactive DNA precursor. HGFH for this assay can be obtained by recombinant means or from biochemical preparations. Incorporation of [3 H]thymidine into acid-precipitable DNA is measured over an appropriate time interval, and the amount incorporated is directly proportional to the amount of newly synthesized DNA. A linear dose-response curve over at least a hundred-fold HGFH concentration range is indicative of growth factor activity. One unit of activity per milliliter is defined as the concentration of HGFH producing a 50% response level, where 100% represents maximal incorporation of [3 H]thymidine into acid-precipitable DNA.

25 XI. Production of HGFH Specific Antibodies

HGFH substantially purified using PAGE electrophoresis (see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols. The HGFH amino acid sequence is analyzed using DNASTAR software (DNASTAR Inc) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well

described in the art. (See, e.g., Ausubel et al. *supra*, ch. 11.)

Typically, the oligopeptides are 15 residues in length, and are synthesized using an Applied Biosystems Peptide Synthesizer Model 431A using fmoc-chemistry and coupled to KLH (Sigma, St. Louis, MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel et al. *supra*.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide activity, for example, by binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

10

XII. Purification of Naturally Occurring HGFH Using Specific Antibodies

Naturally occurring or recombinant HGFH is substantially purified by immunoaffinity chromatography using antibodies specific for HGFH. An immunoaffinity column is constructed by covalently coupling anti-HGFH antibody to an activated chromatographic resin, such as CNBr-activated Sepharose (Pharmacia & Upjohn). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing HGFH are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of HGFH (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/HGFH binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and HGFH is collected.

XIII. Identification of Molecules Which Interact with HGFH

HGFH, or biologically active fragments thereof, are labeled with ¹²⁵I Bolton-Hunter reagent. (See, e.g., Bolton et al. (1973) *Biochem. J.* 133:529.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled HGFH, washed, and any wells with labeled HGFH complex are assayed. Data obtained using different concentrations of HGFH are used to calculate values for the number, affinity, and association of HGFH with the candidate molecules.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with

specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following

5 claims.

What is claimed is:

1. A substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5, and a fragment of SEQ ID NO:7.
2. A substantially purified variant having at least 90% amino acid identity to the amino acid sequence of claim 1.
3. An isolated and purified polynucleotide encoding the HGFH of claim 1.
4. An isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide of claim 3.
5. An isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide of claim 3.
6. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide sequence of claim 3.
7. An isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, a fragment of SEQ ID NO:2, a fragment of SEQ ID NO:4, a fragment of SEQ ID NO:6, and a fragment of SEQ ID NO:8.
8. An isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide of claim 7.
9. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 7.

10. A fragment of the polynucleotide of claim 7 comprising a polynucleotide sequence selected from the group consisting of nucleotides 535-570 of SEQ ID NO:2, nucleotides 124-156 of SEQ ID NO:4, nucleotides 30-62 of SEQ ID NO:6, and nucleotides 147-182 of SEQ ID NO:8.
- 5 11. An expression vector containing at least a fragment of the polynucleotide of claim 3.
12. A host cell containing the expression vector of claim 11.
13. A method for producing a polypeptide comprising the amino acid sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID
10 NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5, or a fragment of SEQ ID NO:7, the method comprising the steps of:
- a) culturing the host cell of claim 12 under conditions suitable for the expression of the polypeptide; and
 - b) recovering the polypeptide from the host cell culture.
- 15 14. A pharmaceutical composition comprising the polypeptide of claim 1 in conjunction with a suitable pharmaceutical carrier.
15. A purified antibody which specifically binds to the polypeptide of claim 1.
16. A purified agonist of the polypeptide of claim 1.
17. A purified antagonist of the polypeptide of claim 1.
- 20 18. A method for treating or preventing a cell proliferative disorder, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 17.
19. A method for treating or preventing a developmental disorder, the method

comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 14.

20. A method for detecting a polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, a fragment of SEQ ID NO:1, a fragment of SEQ ID NO:3, a fragment of SEQ ID NO:5, or a fragment of SEQ ID NO:7 in a biological sample, the method comprising the steps of:

- (a) hybridizing the polynucleotide of claim 6 to at least one of the nucleic acids in the biological sample, thereby forming a hybridization complex; and
- 10 (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of the polynucleotide encoding the polypeptide in the biological sample.

21. The method of claim 20 wherein the nucleic acids of the biological sample are amplified by the polymerase chain reaction prior to hybridization.

1	M	A	R	P	N	K	F	L	L	W	F	C	C	F	A	W	L	C	F	P	I	S	L	G	S	Q	A	S	G	862403	
1	M	A	L	P	S	N	F	L	L	G	V	C	C	F	A	W	L	C	F	L	S	S	L	S	S	Q	A	S	T	E	GI 567206
31	E	A	Q	I	A	A	S	A	E	L	E	S	G	A	M	P	W	S	L	L	Q	H	I	D	E	R	D	R	A	G	862403
31	E	S	Q	S	G	A	S	E	N	V	E	S	E	A	D	P	W	S	L	L	L	P	V	D	G	T	D	R	S	G	GI 567206
61	L	L	P	A	L	F	K	V	L	S	V	G	R	G	G	S	P	R	L	Q	P	D	S	R	A	L	H	Y	M	K	862403
61	L	L	P	P	L	F	K	V	L	S	D	R	R	G	E	T	P	K	L	Q	P	D	S	R	A	L	Y	Y	M	K	GI 567206
91	K	L	Y	K	T	Y	A	T	K	E	G	I	P	K	S	N	R	S	H	L	Y	N	T	V	R	L	F	T	P	C	862403
91	K	L	Y	K	T	Y	A	T	K	E	G	V	P	K	P	S	R	S	H	L	Y	N	T	V	R	L	F	S	P	C	GI 567206
121	T	R	H	K	Q	A	P	G	D	Q	V	T	G	I	L	P	S	V	E	L	L	F	N	L	D	R	I	T	T	V	862403
121	A	Q	Q	E	Q	A	P	S	N	Q	V	T	G	P	L	P	M	V	D	L	L	F	N	L	D	R	V	T	A	M	GI 567206
151	E	H	L	L	K	S	V	L	L	Y	P	L	L	S	K	-	-	-	-	-	-	-	-	-	-	-	-	-	-	862403	
151	E	H	L	L	K	S	V	L	L	Y	T	L	N	N	S	A	S	S	S	S	T	V	T	C	M	C	D	L	V	V	GI 567206
173	K	E	P	K	S	S	R	T	L	G	R	A	P	Y	S	F	T	F	N	S	Q	F	E	F	G	K	K	H	K	862403	
181	K	E	A	M	S	S	G	R	A	P	P	R	A	P	Y	S	F	T	L	-	-	-	-	-	-	-	-	-	-	GI 567206	

FIGURE 1A

203	W	I	Q	I	D	V	T	S	L	L	Q	P	L	V	A	S	N	K	R	S	I	H	M	S	I	N	F	T	C	M	862403
204	W	I	E	I	D	V	T	S	L	L	Q	P	L	V	T	S	S	E	R	S	I	H	L	S	V	N	F	T	C	T	GI 567206
233	K	D	Q	L	E	H	P	S	A	O	N	G	L	E	N	M	T	L	-	V	S	P	S	L	I	L	Y	L	N	D	862403
234	K	D	Q	V	-	-	-	-	P	E	D	G	V	F	S	M	P	L	S	V	P	P	S	L	I	L	Y	L	N	D	GI 567206
262	T	S	A	Q	A	Y	H	S	W	Y	S	L	H	Y	K	R	R	P	S	Q	G	P	D	Q	E	R	S	L	S	A	862403
260	T	S	T	Q	A	Y	H	S	W	Q	S	L	Q	S	T	W	R	P	L	Q	H	P	G	Q	A	-	G	V	A	A	GI 567206
292	Y	P	V	G	E	E	A	A	E	D	G	R	S	S	H	H	R	R	R	G	Q	E	T	V	S	S	E	L	K	862403	
289	R	P	V	K	E	E	A	T	E	V	E	R	S	P	-	-	R	R	R	R	G	Q	K	A	I	R	S	E	A	K	GI 567206
322	K	P	L	G	P	A	S	F	N	L	S	E	Y	F	R	Q	F	L	L	P	Q	N	E	C	E	L	H	D	F	R	862403
317	G	P	L	L	T	A	S	F	N	L	S	E	Y	F	K	Q	F	L	F	P	Q	N	E	C	E	L	H	D	F	R	GI 567206
352	L	S	F	S	Q	L	K	W	D	N	W	I	V	A	P	H	R	Y	N	P	R	Y	C	K	G	D	C	P	R	A	862403
347	L	S	F	S	Q	L	K	W	D	N	W	I	V	A	P	H	R	Y	N	P	R	Y	C	K	G	D	C	P	R	A	GI 567206
382	V	G	H	R	Y	G	S	P	V	H	T	M	V	Q	N	I	I	Y	E	K	L	D	S	S	V	P	R	P	S	C	862403
377	V	R	H	R	Y	G	S	P	V	H	T	M	V	Q	N	I	I	Y	E	K	L	D	P	S	V	P	R	P	S	C	GI 567206

FIGURE 1B

412	V P A K Y S P L S V L T I E P D G S I A Y K E Y E D M I A T	862403
407	V P G K Y S P L S V L T I E P D G S I A Y K E Y E D M I A T	GI 567206
442	K C T C R	862403
437	R C T C R	GI 567206

FIGURE 1C

1 M A R P - R P R E Y K A G D L V F A K M K G Y P H W P A R I 2676869
 1 M S R S N R Q K E Y K C G D L V F A K M K G Y P H W P A R I GI 598956

 30 D E L P E G A V K P P A N K Y P I F F F G T H E T A F L G P 2676869
 31 D E M P E A A V K S T A N K Y Q V F F F G T H E T A F L G P GI 598956

 60 K D L F P Y K E Y K D K F G K S N K R K G F N E G L W E I E 2676869
 61 K D L F P Y E E S K E K F G K P N K R K G F S E G L W E I E GI 598956

 90 N N P G V K F T G Y Q A I Q Q S S S E - - - - - T E 2676869
 91 N N P T V K A S G Y Q S S Q K K S C V E E P E P E A A E GI 598956

 112 G E G G N T A D A - S S E E E G D R V E E D G K G K R K N E 2676869
 121 G D G D K K G N A E G S S D E E G K L V I D E P A K E K N E GI 598956

 141 K A G S K R K K S Y - - - T S K K S S K Q S R K S P G D E D 2676869
 151 K G A L K R R A G D L L E D S P K R P K E A E N P E G E E K GI 598956

 168 D K D C K E - - - - - E E N K S S S E G G D A G N D 2676869
 181 E A A T L E V E R P L P M E V E K N S T P S E P G S G R G P GI 598956

FIGURE 2A

189 TRN TTS D L Q K T S E G T 2676869
 211 P Q E E E E E D E E E E A T K E D A E A P G I R D H E S L G I 598956

FIGURE 2B

1	MGRKVT	VATCA	LNQW	ALDF	EGNL	ORIL	KS	I	1568019
1	MSHLIT	LATCN	LNQW	ALDF	EGNR	DRIL	QS	I	GI 500832
31	EIAK	NRGAR	YRLG	PELE	ICGY	GCWD	HY	ES	1568019
31	KIAK	ERGAR	LRVG	PELE	ITGY	GC	LDH	FL	EN
61	DTLL	HSFQ	VLA	ALLES	PVTQ	DIIC	DV	GM	PV
61	DVCL	HSWEM	YAQ	IIK	NKET	HGLI	LDI	G	MP
91	MHRN	VRYN	CRVI	FLNR	KILL	IRPK	MA	LA	NE
91	LHK	NVRY	NCRL	LSLD	GEIL	FIRPK	IW	LA	ND
121	GN	YREL	RWF	TPWS	RSRH	TEEY	FL	PR	MI
121	GN	YRE	MRFF	TPWM	KPGV	VEDFI	LP	PEI	QK
151	TKQ	ETVP	FFGD	AVLV	TWDT	CTCI	GS	EICE	EL
151	TGQ	RLVP	FFGD	AVIN	SLDT	CTCI	GTET	CEEL	FT
181	PH	SPHI	IDM	GLDG	VEIIT	NASG	SHH	VLR	KAN
181	PQ	SPHI	AM	SLDG	VEIM	TNSG	SHH	ELR	KLN

FIGURE 3A

211	T R V D L V T M V T S K N G G I Y L L A N Q K G C D G D R L	1568019
211	K R L D L I L N A T K R C G G V Y L Y A N Q R G C D G D R L	GI 500832
241	Y Y D G C A M I A M N G S V F A Q G S Q F S L D D V E V L T	1568019
241	Y Y D G C A L I A I N G T I V A Q G S Q F S L D D V E V T	GI 500832
271	A T L D L E D V R S Y R A E I S S R N L A A S R A S - P Y P	1568019
271	A T V D L E E V R S Y R A A V M S R G L Q A S L A E I K F K	GI 500832
300	R V K V D F A L S C H E D L L A P I S E P I E W K - - Y H	1568019
301	R I D I P V E L A L M T S R F D P T V C P T K V R E P F Y H	GI 500832
327	S P E E E I S L G P A C W L W D F L R R S Q Q A G F L L P L	1568019
331	S P E E E I A L G P A C W M W D Y L R R C N G T G F F L P L	GI 500832
357	S G G V D S A A T A C L I Y S M C C Q V C E A V R S G N E E	1568019
361	S G G I D S C A T A M I V H S M C R L V T D A A Q N G N E Q	GI 500832
387	V L A D V R T I V - N Q I S Y T P Q D P R D L C G R I L T T	1568019
391	V I K D V R K I T R S G D D W I P D S P Q D L A S K I F H S	GI 500832

FIGURE 3B

416	CYMASKNSSQ	ETCTRA	RELA	QQIG	SHHISL	1568019
421	CFMG	TENSSKET	RNR	AKDL	SNAL	GI 500832
					YHVDL	
446	NIDPAVK	AVMGI	FSLV	TGKS	PLFA	1568019
451	KMD	SLVSSV	VSLFE	VATG	KKPI	GI 500832
					YKIF	
476	REN	LALQN	VQAR	IRMV	LA	1568019
481	IEN	LALQN	IQAR	LRMV	LSY	GI 500832
					LFAL	
					LPV	
506	VH	-	-	GG	LLVL	1568019
511	IPNS	GG	LLVL	GSAN	VDE	GI 500832
					CLRG	
534	ADIN	PIGG	ISKT	DLRA	FVQ	1568019
541	ADIN	PIGG	ISKT	DLKR	FI	GI 500832
					AYAS	
					KQYN	
					MPIL	
564	OS	ILL	AP	ATA	ELE	1568019
571	ND	FL	NA	TP	TA	GI 500832
					ELE	
					PM	
					TKDY	
					V-Q	
					SD	
					EID	
594	TY	AEL	S	VY	GK	1568019
600	TY	EEL	GV	FGY	LR	GI 500832
					KKV	
					AKM	
					G	
					GPYS	
					MF	
					CKLL	
					LG	
					MLH	
					QW	

FIGURE 3C

9/11

624	R	H	I	C	T	P	R	Q	V	A	D	K	V	K	R	F	F	S	K	Y	S	M	N	R	H	K	M	T	T	L	1568019
630	S	P	K	L	T	P	R	Q	I	S	E	K	V	K	R	F	F	F	F	Y	A	I	N	R	H	K	Q	T	V	L	GI 500832
654	T	P	A	Y	H	A	E	N	Y	S	P	E	D	N	R	F	D	L	R	P	F	L	Y	N	T	S	W	P	W	Q	1568019
660	T	P	S	Y	H	A	E	Q	Y	S	P	E	D	N	R	F	D	L	R	P	F	L	I	N	P	R	F	P	W	A	GI 500832
684	F	R	C	I	E	N	Q	V	L	Q	L	E	R	A	E	P	Q	S	L	D	-	-	G	V	D						1568019
690	S	R	K	I	D	E	V	V	E	Q	C	E	A	H	K	G	S	T	L	D	I	M	S	I	D						GI 500832

FIGURE 3D

1 MKLH - YVAVLT LA ILMFLT W - - LPESLSCN 3577857
1 MQLLCYFVILFVGI - - - APWSS LANDDGCN GI 529900

28 KALCASDVSKCLIQEELCQCRPGEGNCSCCK 3577857
28 EVVCGSVVSCKCLITQSCQCKLND - - CHCCK GI 529900

58 ECLCLGALWD ECCCDCVGM CNPRNYS DTPP 3577857
56 DCLNCLGELYIECCGC LDMC PKHKDVLPSL GI 529900

88 TSKSTVEELHEPIPSLFRALTEGDTQLNWN 3577857
86 TPRS EIGDI - EGVPELFDTLTAE DDE - GWS GI 529900

118 IVSFPVAEELSHHENLVSFLETVNQPHHQN 3577857
114 TIRFSMRA - - - - - GFKQRVQGGASGD GI 529900

148 VSVPSNNVHAPYSSSDKEHMC TVVYFDDCMS 3577857
135 AGNGNGNA - - - GSAGVTLCTVIYVNSCIR GI 529900

178 IHQCKISCESMGASKYRWFHNA CCECIGPE 3577857
163 ANKCRQCCESMGASSYRWFHDG CCECVGEN GI 529900

FIGURE 4A

208	C	I	D	Y	G	S	K	T	V	K	C	M	N	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3577857
193	C	L	N	Y	G	I	N	E	S	R	C	R	G	C	P	E	D	Q	D	Q	L	L	T	A	D	T	V	P	A	E	GI 529900
222	-	-	-	-	-	-	-	-	-	-	-	-	-	-	M	F															3577857
223	A	E	Q	D	L	E	R	F	F	G	N	E	E	I	E	D	E	W	G	Y	G	E	E	D	E	F	S	GI 529900			

FIGURE 4B

SEQUENCE LISTING

<110> INCYTE PHARMACEUTICALS, INC.
 HILLMAN, Jennifer L.
 LAL, Preeti
 CORLEY, Neil C.
 BAUGHN, Mariah
 GUEGLER, Karl J.
 TANG, Y. Tom

<120> HUMAN GROWTH FACTOR HOMOLOGS

<130> PF-0472 PCT

<140> To Be Assigned

<141> Herewith

<150> 09/015,412

<151> 1998-01-29

<160> 12

<170> PERL PROGRAM

<210> 1

<211> 446

<212> PRT

<213> Homo sapiens

<220> -

<223> 862403

<400> 1

Met	Ala	Arg	Pro	Asn	Lys	Phe	Leu	Leu	Trp	Phe	Cys	Cys	Phe	Ala
1				5					10					15
Trp	Leu	Cys	Phe		Ile	Ser	Leu	Gly	Ser	Gln	Ala	Ser	Gly	Gly
				20					25					30
Glu	Ala	Gln	Ile	Ala	Ala	Ser	Ala	Glu	Leu	Glu	Ser	Gly	Ala	Met
				35					40					45
Pro	Trp	Ser	Leu	Leu	Gln	His	Ile	Asp	Glu	Arg	Asp	Arg	Ala	Gly
				50					55					60
Leu	Leu	Pro	Ala	Leu	Phe	Lys	Val	Leu	Ser	Val	Gly	Arg	Gly	Gly
				65					70					75
Ser	Pro	Arg	Leu	Gln	Pro	Asp	Ser	Arg	Ala	Leu	His	Tyr	Met	Lys
				80					85					90
Lys	Leu	Tyr	Lys	Thr	Tyr	Ala	Thr	Lys	Glu	Gly	Ile	Pro	Lys	Ser
				95					100					105
Asn	Arg	Ser	His	Leu	Tyr	Asn	Thr	Val	Arg	Leu	Phe	Thr	Pro	Cys
				110					115					120
Thr	Arg	His	Lys	Gln	Ala	Pro	Gly	Asp	Gln	Val	Thr	Gly	Ile	Leu
				125					130					135
Pro	Ser	Val	Glu	Leu	Leu	Phe	Asn	Leu	Asp	Arg	Ile	Thr	Thr	Val
				140					145					150
Glu	His	Leu	Leu	Lys	Ser	Val	Leu	Leu	Tyr	Pro	Leu	Leu	Ser	Lys
				155					160					165
Cys	Val	Cys	Asn	Leu	Met	Ile	Lys	Glu	Pro	Lys	Ser	Ser	Ser	Arg
				170					175					180

Thr	Leu	Gly	Arg	Ala	Pro	Tyr	Ser	Phe	Thr	Phe	Asn	Ser	Gln	Phe	185	190	195
Glu	Phe	Gly	Lys	Lys	His	Lys	Trp	Ile	Gln	Ile	Asp	Val	Thr	Ser	200	205	210
Leu	Leu	Gln	Pro	Leu	Val	Ala	Ser	Asn	Lys	Arg	Ser	Ile	His	Met	215	220	225
Ser	Ile	Asn	Phe	Thr	Cys	Met	Lys	Asp	Gln	Leu	Glu	His	Pro	Ser	230	235	240
Ala	Gln	Asn	Gly	Leu	Phe	Asn	Met	Thr	Leu	Val	Ser	Pro	Ser	Leu	245	250	255
Ile	Leu	Tyr	Leu	Asn	Asp	Thr	Ser	Ala	Gln	Ala	Tyr	His	Ser	Trp	260	265	270
Tyr	Ser	Leu	His	Tyr	Lys	Arg	Arg	Pro	Ser	Gln	Gly	Pro	Asp	Gln	275	280	285
Glu	Arg	Ser	Leu	Ser	Ala	Tyr	Pro	Val	Gly	Glu	Glu	Ala	Ala	Glu	290	295	300
Asp	Gly	Arg	Ser	Ser	His	His	Arg	His	Arg	Arg	Gly	Gln	Glu	Thr	305	310	315
Val	Ser	Ser	Glu	Leu	Lys	Lys	Pro	Leu	Gly	Pro	Ala	Ser	Phe	Asn	320	325	330
Leu	Ser	Glu	Tyr	Phe	Arg	Gln	Phe	Leu	Leu	Pro	Gln	Asn	Glu	Cys	335	340	345
Glu	Leu	His	Asp	Phe	Arg	Leu	Ser	Phe	Ser	Gln	Leu	Lys	Trp	Asp	350	355	360
Asn	Trp	Ile	Val	Ala	Pro	His	Arg	Tyr	Asn	Pro	Arg	Tyr	Cys	Lys	365	370	375
Gly	Asp	Cys	Pro	Arg	Ala	Val	Gly	His	Arg	Tyr	Gly	Ser	Pro	Val	380	385	390
His	Thr	Met	Val	Gln	Asn	Ile	Ile	Tyr	Glu	Lys	Leu	Asp	Ser	Ser	395	400	405
Val	Pro	Arg	Pro	Ser	Cys	Val	Pro	Ala	Lys	Tyr	Ser	Pro	Leu	Ser	410	415	420
Val	Leu	Thr	Ile	Glu	Pro	Asp	Gly	Ser	Ile	Ala	Tyr	Lys	Glu	Tyr	425	430	435
Glu	Asp	Met	Ile	Ala	Thr	Lys	Cys	Thr	Cys	Arg					440	445	

<210> 2

<211> 2592

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> 135, 2396, (2402)...(2412), (2417)...(2422),

<221> unsure

<222> (2429)...(2432), 2434, 2436, 2441, 2442, 2444

<221> unsure

<222>

<223> a o r g o r c o r t, unknown, or other

<220> -

<223> 862403

<400> 2

```

ggggtagata cgcggtccggt ggggcataaa caggaatttg cttgtcgggt tcgtactcag 60
gaacgaacat cttaggettc aatttctcaa aaaatgagat attttattta agatcattcc 120
tcaattttatc ttagnacttt cagctacata gttagtcggg tagtgtaaca gtaacggggt 180
ttacgtaaat gccaaaggaa aagagaggaa agctaaataa ccaagtaaat ctgtatctaa 240
ttaacaaatt ggctaaataa ggtgttatca gctgcttgat atagagctga taaaatcttc 300
agctaggcat acttgaggcc tgattacaga agtgaccgta gtccaccac acacctgaaa 360
tttatttaag aagaccaagc taggctcttc ctggccttta ggaagaggac tggcatggag 420
aaatatgttc ctactagtt ctcccaagcc atggcacgtc ccaacaaatt cctcctttgg 480
ttttgctgct ttgctgggt gtgttttccct attagccttg gttctcaggc ttctggggga 540
gaagctcaga ttgctgctag tgctgagttg gaatctgggg ctatgccttg gtccttgctg 600
cagcatatag atgagagaga cagagctggc ctccctcccg cgcttttcaa agttctatct 660
gttggggcgag gttgggtcacc taggctgcag ccagactcca gagctttgca ctacatgaag 720
aagctctata agacatatgc taccaaggaa gggattccta aatccaatag aagtcacctc 780
tacaacactg ttcggtctct caccctctgt acccggcaca agcaggctcc tggagaccag 840
gtaacaggaa tccttccatc agtggaaact ctatttaacc tggatcgcat tactaccgtt 900
gaacacttac tcaagtcagt cttgctgtac cctctgctgt ccaaattgtg gtgcaatcta 960
atgataaagg agccaaagtc ttctagcagg actctcgga gagctccata ctcatctacc 1020
tttaactcac agtttgaatt tggaaagaaa cacaatgga ttcagattga tgtgaccagc 1080
ctccttcaac ctttagtggc ctccaacaag agaagtattc acatgtctat aaattttact 1140
tgcatgaaag accagctgga gcatccttca gcacagaatg gtttgtttaa catgactctg 1200
gtgtccccct cactgatctt atatttgaat gacacaagt ctcaggctta tcacagctgg 1260
tattcccttc actataaaa gaggccttcc cagggtcctg accaggagag aagtctgtct 1320
gcctatcctg tgggagaaga ggctgctgag gatgggagat cttcccatca ccgtcaccgc 1380
agaggtcagg aaactgtcag ttctgaattg aagaagccct tgggcccagc ttccttcaat 1440
ctgagtgaat acttcagaca atttcttctt ccccaaatg agtggtgagct ccatgacttt 1500
agacttagct ttagtcagct gaagtgggac aactggattg tggctccgca caggtaacaac 1560
cctcgatact gtaaaggggg ctgtccaagg gcagttggac atcggtatgg ctctccagtt 1620
cacaccatgg tacagaacat catctatgag aagctggact cctcagtgcc aagaccgtca 1680
tgtgtacctg ccaaatacag ccccttgagt gttttgacca ttgagcccga tggctcaatt 1740
gcctataaag agtacgaaga tatgatagct acaaagtgca cctgtcgtta acaaattggtc 1800
ctcttaaaac cttgagccta tttggcaaag taactactgt gtgcctatgt gtgccttcaa 1860
gagaaagctt catatattaa gtctctaaat gtagcatatg ttatataagg aggagcctgt 1920
gtaggttagt acctctatg gcatctatca ggataaaggg ataacatcaa ttgttgctac 1980
agagcctttt ttttatttcc aaattttaat gaaatataat tattgtggag aactttacat 2040
ttttttcctt gagtgatatt ttttcttttc ataggagtct tattcttgat aggggaaaaa 2100
ccttaattag catcaatcct ggatggactt gcagctataa ataggcaatt cagattgctg 2160
tagtcttaat agaagaataa atttctgtca atggcaaaaa aaaaaaaaaa 2220
aaaagggggg gccctcaaa gggccccaag ttttgggagc cggggggttg ggggtaaaac 2280
ccccttgaag gggcccccac atttcatttc cggggggggg tttttaaaag ggggggttgg 2340
ggaaaccccg ggggttcccc attttaggcc gcttcagggc ccccccccaa aggggnacgg 2400
gnnnnnnnnn nnaattnnnn nntcctttnn nntnncctc nncncccttg aagttgcatt 2460
tggttcttag tggccggatt tagagattca ggacttttcc tgtttggcca ggttgatgct 2520
attgttgtgt ccttggtgga ttccaccag gtttgtgctg ttgtgggcgt gagtgtcccc 2580
ggttgtcgcg gt 2592

```

<210> 3

<211> 203

<212> PRT

<213> Homo sapiens

<220> -

<223> 2676869

<400> 3

Met Ala Arg Pro Arg Pro Arg Glu Tyr Lys Ala Gly Asp Leu Val

1	5	10	15
Phe Ala Lys Met	Lys Gly Tyr Pro His	Trp Pro Ala Arg	Ile Asp
20	25	30	
Glu Leu Pro Glu	Gly Ala Val Lys Pro	Pro Ala Asn Lys	Tyr Pro
35	40	45	
Ile Phe Phe Phe	Gly Thr His Glu Thr	Ala Phe Leu Gly	Pro Lys
50	55	60	
Asp Leu Phe Pro	Tyr Lys Glu Tyr Lys	Asp Lys Phe Gly	Lys Ser
65	70	75	
Asn Lys Arg Lys	Gly Phe Asn Glu Gly	Leu Trp Glu Ile	Glu Asn
80	85	90	
Asn Pro Gly Val	Lys Phe Thr Gly Tyr	Gln Ala Ile Gln	Gln Gln
95	100	105	
Ser Ser Ser Glu	Thr Glu Gly Glu Gly	Gly Asn Thr Ala	Asp Ala
110	115	120	
Ser Ser Glu Glu	Glu Gly Asp Arg Val	Glu Glu Asp Gly	Lys Gly
125	130	135	
Lys Arg Lys Asn	Glu Lys Ala Gly Ser	Lys Arg Lys Lys	Ser Tyr
140	145	150	
Thr Ser Lys Lys	Ser Ser Lys Gln Ser	Arg Lys Ser Pro	Gly Asp
155	160	165	
Glu Asp Asp Lys	Asp Cys Lys Glu Glu	Glu Asn Lys Ser	Ser Ser
170	175	180	
Glu Gly Gly Asp	Ala Gly Asn Asp Thr	Arg Asn Thr Thr	Ser Asp
185	190	195	
Leu Gln Lys Thr	Ser Glu Gly Thr		
200			

<210> 4

<211> 3805

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> 3788

<223> a or g or c or t, unknown, or other

<220> -

<223> 2676869

<400> 4

```

gcgaccgctc gtccgcgggc ttgaggcccg cggggaggcg cgcaattcgt cggccccggg 60
gggggaggcc atcccggcag tcttcgcggc gaccaaggac taccaggaag gggagcggct 120
gggatggcgc gtccgcgggc ccgcgagtag aaagcgggag acctgggtctt cgccaagatg 180
aagggtacc cgcactggcc ggcccggatt gatgaactcc cagagggcgc tgtgaagcct 240
ccagcaaaca agtatcctat cttctttttt ggcacccatg aaactgcatt tctaggtccc 300
aaagaccttt ttccatataa ggagtacaaa gacaagtttg gaaagtcaaa caaacggaaa 360
ggatttaacg aaggattgtg ggaaatagaa aataaccag gagtaaagtt tactggctac 420
caggcaattc agcaacagag ctcttcagaa actgaggag aaggtggaaa tactgcagat 480
gcaagcagtg aggaagaagg tgatagagta gaagaagatg gaaaaggcaa aagaaagaat 540
gaaaaagcag gctcaaaacg gaaaaagtca tatacttcaa agaaatcctc taaacagtcc 600
cggaaatctc caggagatga agatgacaaa gactgcaaag aagaggaaaa caaaagcagc 660
tctgaggggtg gagatgcggg caacgacaca agaaacacaa cttcagactt gcagaaaacc 720

```

```

agtgaaggga cctaactacc ataatgaatg ctgcatatta agagaaacca caagaaggtt 780
atatgtttgg ttgtctaata ttcttggatt tgatatgaac caacacatag tccttgttgt 840
cattgacaga accccagttt gtatgtacat ttttcatatt cctctctgtt gtgtttcggg 900
gggaaaagac attttagcct tttttaaaag ttactgattt aatttcatgt tatttggttg 960
catgaagttg cccttaacca ctaaggatta tcaagatttt tgcgcagact tatacatgtc 1020
taggatccctt ttatcaaggc agttatgata atcgttttcc tgccttgacc ccaccatcat 1080
caaacactca gttaaatata aattaacatt ttttagatga ccactcaaca taatgcttaa 1140
gaatggaatt tcctctctgt gacagaaccc aggaattaat tcctaaatag ataacgttgg 1200
tatattgaag acgaaattaa aattgtccct cagttttgag gccatgtgta aagtttaacc 1260
atatgttaaa atatctattc cgtattagaa atagctagtt gacagcttat acttctcaaa 1320
attcatattg ttatgtacac aaactaagtt tctatatgtg aagttagtga gtctttttgt 1380
gttactccaa aataaaggca atgattttatt tttttcccag tgccaataca attttgagct 1440
aagcactcaa ggtggatact ttacatttta aagctggaat cagcaacagc cctatgggaa 1500
accagacaaa gcattgactt ttaaatgtag acttttaaaa taaactgttt tcttttggaa 1560
ctacaattag aatagttaat attcatcctt aaaccattat tatgtgtaca ttattgttgc 1620
tattgtgata atagagaatt ttattttatt ttatgccagc ttatattgtg agaacacatt 1680
tagtcagttt ggggtttatc aatcctgtta atgcttgtcc ttggaacatc tttcgcgtat 1740
tcacggtttg tagttgaaaa gtttactgta aaaaaatcaa aaacaaaaaa atgtattgtt 1800
tttacagaat aaatttattg gaatgtgtac tgggagtaag atttgaggtt gtaagcaaac 1860
taagtttagt taatttggct tcatatatgt aacgtgaggt ataatgttaa ttcatatatt 1920
aaagcaaaaa ttgttcacag caagctgaca atagaatcaa gtgcaggtga gggtttttct 1980
ttttctttct tttttttttt ttaaacactc tctgggtttt agattatttg aaaacactgt 2040
aggggatgat ggggagatga ctacatgatt tgcttacct tacatattta ccagcccagt 2100
ggaccctaga ttttgaatt tgatattatc catgcattga tttgtatccc attttctttt 2160
ggttttcttg atgagagcta ataaaaacat gagaaaatgt gagcagacct aaaagcctag 2220
ggccgttaat aatgatgatg gtgatgatgg cagctaaggg tatgaaagca ccagtcattt 2280
tatcattcca gggcacagtg aggaggtggc tgtttgtatt tttggtgtgt tcttttcatt 2340
tagtgaatac tgattttctc tagtatactc aacatacaat cccaacatg aaaagttcat 2400
aaaaagatag tagggatgtc gatctaatac ttcttcctta aaaggctcct ttttaattttg 2460
aatatttgaa tgtatttttt aattaatgaa atttaattca tcaaaagacc tcttcattgc 2520
tcacactgaa acataagggg aatgtagatg gcattttgaa ttggaatata caactcaca 2580
attttcttta aaagcgtttt atgtgtatat atgaaatggt tcaaatttaa taagatttgt 2640
ctttataaga ggccaaaaat aattttatta aaaatataca cagaaatata atagaaatat 2700
ttccagccac caataaaaca aatgttcatt aaggatagag ataaagttac tgatccctat 2760
ttagctgttt aaattcacac aaaaaaattt agtaaaagta attttcttga aacagcagca 2820
aataataaac tttccttggg aaaacatttt tcaaaacaac tttaaaaata ttctgatgat 2880
ttgcctttta tattgagttt aataatgaag catattcatc ttatttgcaa aagttactcc 2940
tccagtctgc attcactgtt ttgaattttt taatatgcac aggccatatt ctcttttgaa 3000
tagtttcatt tcatgctcac tactgttggc actaaattgt aattttttac acttataatt 3060
attttatatg tcaacaata ttgagatgga tgatcaagtt tattcacaata tgattttgaa 3120
tatatttctt taaacaaact gtttctcaag aatttccatt ttccttcaaa aaaaatttac 3180
acaataccat atagctcaca catacagtgc agagcaagaa tgtacatatt ctttaagtact 3240
cacgtatgag aagagagtga aatgggacaa aatatcagta ccagtataac aaaagttagt 3300
aatttagaaa ggaggaacat gaagtaatat ttttattcca ctttttcttc tgctttgtt 3360
tttatgaaga actctggttt gtagatacaa cgataggcca agagctgagg aagaactcca 3420
tatgctatgt ttaatgctaa aaaaaggatt tttgcttctt cagggactct gtagacataa 3480
gcagttctag catgaagaga tgcaccaatg tgagaaaact gagcctgagc cagacctcca 3540
gcagtatca atgtgatgtc aggcattccag gaacatccag gaaccactaa gccatacagt 3600
gcagtcacaa agtaaggaa acgaatagaac atatatgcca gcatctgaat tttaggataa 3660
gcagcaggat cctttagata gggctcttga aattgcgtat ataatcgga gagctcagat 3720
gggcaatcca aagcaatcaa acctctgaac aggcaaaatc cagttgcccg gagcgtggg 3780
cccacgcntt ccgcccacgc gtctg 3805

```

<210> 5

<211> 706

<212> PRT

<213> Homo sapiens

<220> -

<223> 1568019

<400> 5

```

Met Gly Arg Lys Val Thr Val Ala Thr Cys Ala Leu Asn Gln Trp
  1          5          10          15
Ala Leu Asp Phe Glu Gly Asn Leu Gln Arg Ile Leu Lys Ser Ile
          20          25          30
Glu Ile Ala Lys Asn Arg Gly Ala Arg Tyr Arg Leu Gly Pro Glu
          35          40          45
Leu Glu Ile Cys Gly Tyr Gly Cys Trp Asp His Tyr Tyr Glu Ser
          50          55          60
Asp Thr Leu Leu His Ser Phe Gln Val Leu Ala Ala Leu Leu Glu
          65          70          75
Ser Pro Val Thr Gln Asp Ile Ile Cys Asp Val Gly Met Pro Val
          80          85          90
Met His Arg Asn Val Arg Tyr Asn Cys Arg Val Ile Phe Leu Asn
          95          100          105
Arg Lys Ile Leu Leu Ile Arg Pro Lys Met Ala Leu Ala Asn Glu
          110          115          120
Gly Asn Tyr Arg Glu Leu Arg Trp Phe Thr Pro Trp Ser Arg Ser
          125          130          135
Arg His Thr Glu Glu Tyr Phe Leu Pro Arg Met Ile Gln Asp Leu
          140          145          150
Thr Lys Gln Glu Thr Val Pro Phe Gly Asp Ala Val Leu Val Thr
          155          160          165
Trp Asp Thr Cys Ile Gly Ser Glu Ile Cys Glu Glu Leu Trp Thr
          170          175          180
Pro His Ser Pro His Ile Asp Met Gly Leu Asp Gly Val Glu Ile
          185          190          195
Ile Thr Asn Ala Ser Gly Ser His His Val Leu Arg Lys Ala Asn
          200          205          210
Thr Arg Val Asp Leu Val Thr Met Val Thr Ser Lys Asn Gly Gly
          215          220          225
Ile Tyr Leu Leu Ala Asn Gln Lys Gly Cys Asp Gly Asp Arg Leu
          230          235          240
Tyr Tyr Asp Gly Cys Ala Met Ile Ala Met Asn Gly Ser Val Phe
          245          250          255
Ala Gln Gly Ser Gln Phe Ser Leu Asp Asp Val Glu Val Leu Thr
          260          265          270
Ala Thr Leu Asp Leu Glu Asp Val Arg Ser Tyr Arg Ala Glu Ile
          275          280          285
Ser Ser Arg Asn Leu Ala Ala Ser Arg Ala Ser Pro Tyr Pro Arg
          290          295          300
Val Lys Val Asp Phe Ala Leu Ser Cys His Glu Asp Leu Leu Ala
          305          310          315
Pro Ile Ser Glu Pro Ile Glu Trp Lys Tyr His Ser Pro Glu Glu
          320          325          330
Glu Ile Ser Leu Gly Pro Ala Cys Trp Leu Trp Asp Phe Leu Arg
          335          340          345
Arg Ser Gln Gln Ala Gly Phe Leu Leu Pro Leu Ser Gly Gly Val
          350          355          360
Asp Ser Ala Ala Thr Ala Cys Leu Ile Tyr Ser Met Cys Cys Gln
          365          370          375
Val Cys Glu Ala Val Arg Ser Gly Asn Glu Glu Val Leu Ala Asp

```

Val Arg Thr Ile	380	Val Asn Gln Ile Ser Tyr Thr Pro Gln Asp Pro	385	390
	395		400	405
Arg Asp Leu Cys Gly Arg Ile Leu Thr Thr Cys Tyr Met Ala Ser	410		415	420
Lys Asn Ser Ser Gln Glu Thr Cys Thr Arg Ala Arg Glu Leu Ala	425		430	435
Gln Gln Ile Gly Ser His His Ile Ser Leu Asn Ile Asp Pro Ala	440		445	450
Val Lys Ala Val Met Gly Ile Phe Ser Leu Val Thr Gly Lys Ser	455		460	465
Pro Leu Phe Ala Ala His Gly Gly Ser Ser Arg Glu Asn Leu Ala	470		475	480
Leu Gln Asn Val Gln Ala Arg Ile Arg Met Val Leu Ala Tyr Leu	485		490	495
Phe Ala Gln Leu Ser Leu Trp Ser Arg Gly Val His Gly Gly Leu	500		505	510
Leu Val Leu Gly Ser Ala Asn Val Asp Glu Ser Leu Leu Gly Tyr	515		520	525
Leu Thr Lys Tyr Asp Cys Ser Ser Ala Asp Ile Asn Pro Ile Gly	530		535	540
Gly Ile Ser Lys Thr Asp Leu Arg Ala Phe Val Gln Phe Cys Ile	545		550	555
Gln Arg Phe Gln Leu Pro Ala Leu Gln Ser Ile Leu Leu Ala Pro	560		565	570
Ala Thr Ala Glu Leu Glu Pro Leu Ala Asp Gly Gln Val Ser Gln	575		580	585
Thr Asp Glu Glu Asp Met Gly Met Thr Tyr Ala Glu Leu Ser Val	590		595	600
Tyr Gly Lys Leu Arg Lys Val Ala Lys Met Gly Pro Tyr Ser Met	605		610	615
Phe Cys Lys Leu Leu Gly Met Trp Arg His Ile Cys Thr Pro Arg	620		625	630
Gln Val Ala Asp Lys Val Lys Arg Phe Phe Ser Lys Tyr Ser Met	635		640	645
Asn Arg His Lys Met Thr Thr Leu Thr Pro Ala Tyr His Ala Glu	650		655	660
Asn Tyr Ser Pro Glu Asp Asn Arg Phe Asp Leu Arg Pro Phe Leu	665		670	675
Tyr Asn Thr Ser Trp Pro Trp Gln Phe Arg Cys Ile Glu Asn Gln	680		685	690
Val Leu Gln Leu Glu Arg Ala Glu Pro Gln Ser Leu Asp Gly Val	695		700	705
Asp				

<210> 6

<211> 2762

<212> DNA

<213> Homo sapiens

<220> -

<223> 1568019

<400> 6

ggcctcctgc ccaagcgact gcggccagga tgggccggaa ggtgaccgtg gccacctgcg 60
cactcaacca gtggggcctg gacttcgagg gcaatttgca aagaatttta aagagtattg 120
aaattgccaa aaacagagga gcaagataca ggcttggacc agagctggaa atatgctgct 180
acggatgttg ggatcattat tacgagtcgg acaccctctt gcactcgttt caagtcctag 240
cgcccttctt ggagtctccc gtcactcagg acatcatctg cgacgtgggg atgcctgtaa 300
tgcaccgaaa cgtccgctac aactgcagag tgatattcct caacaggaag atcctgctca 360
tcagacccaa gatggccttg gccaatgaag gcaactaccg cgagctgcgc tggttcacc 420
cgtggtcgag gagtccggac acagaggagt actttctgcc tcggatgata caggacctga 480
caaagcagga aaccgtaccc ttccgagatg cgtgctggt gacatgggac acctgcattg 540
gaagtggat ctgtgaggag ctctggacac cccacagccc gcacatcgac atgggcctg 600
atggcgtgga gatcatcacc aacgcctcgg gcagccacca cgtgctgcgc aaagccaaca 660
ccagggtgga tctcgtgact atggtcacca gcaagaacgg tgggatttac ttgctggcca 720
accagaaggg ttgacgagg gaccgcctgt actacgacgg ctgtgccatg atcgccatga 780
acggaagcgt ctttgctcaa ggatcccagt tttctctgga tgacgtggaa gtcctgacgg 840
ccacgctgga tctggaggac gtccggagct acagggcgga gatttcatct cgaaacctg 900
cgccagcag ggagagcccc taccacagag tgaagtgga ctttgccctc tcgtgccacg 960
aggacttgct ggcacccatc tctgagccca tcgagtggaa ataccacagc cctgaggagg 1020
agataagcct tggacctgcc tgctggctct gggatttttt aagacgaagt caacaggcag 1080
ggtttttgct gcccttgagt ggcggggtgg acagcgagc caccgcctgc ctcatctact 1140
ccatgtgctg ccaggtctgc gaggccgtga ggagtggaaa tgagggaagt ctggctgatg 1200
tccgcaccat cgtgaaccag atcagctaca cccccagga tccccgagac ctctgtggac 1260
gcatactgac cactgctac atggccagca agaactcctc ccaggagacg tgcaccggg 1320
ccagagagtt ggcccagcag attggaagcc accacatcag tctcaacatc gatccagcg 1380
tgaagggcgt catgggcatc ttcagcctgg tgacggggaa gagccctctg tttgcagctc 1440
atggaggaa cagcagggaa aacctggcg tgcaaaatgt gcaggctcga atacggatgg 1500
tcctcgcta tctgtttgct cagttgagcc tctggtctcg ggtgtccac ggtgggctcc 1560
tcgtgctggg atccgccaac gtggatgaga gtctcctggg ctacctgacc aagtacgact 1620
gtccagtgcc ggacatcaac cccataggcg ggatcagcaa gacggacctc agggccttcg 1680
tccagtctg catccagcgc ttccagcttc ctgccctgca gagcatcctg ttggcgccg 1740
ccaccgaga gctggagccc ttggctgatg gacaggtgtc ccagaccgac gaggaagata 1800
tgggatgac atatgcggag ctctcgtct atgggaaact caggaagggt gccaaagatg 1860
ggccctacag catgttctgc aaactcctcg gcatgtggag acacatctgc accccgagac 1920
aggctcgtga caaagtgaag cggtttttct ccaagtactc catgaacaga cacaagatga 1980
ccacgctcac acccgctac cagccgaga actacagccc tgaggacaac aggtttgatc 2040
tgccaccatt tctgtacaac acaagctggc cttggcagtt tcggtgcata gaaaatcagg 2100
tgctacagct cgagagggca gagccacagt ccctggacgg cgtggactga ggccggttcc 2160
ttcctggagg cctcctgtcc tcggggaccc cagcacctca tcatcagcat tgctggagcc 2220
aagggtagga gccctacact aggagcccag gatgggacgg cgcacagcc gagagggagg 2280
gaacttttca gtcaaattcc tcaaaaagag gctggaataa agcctgggct taaaaagagg 2340
ctggaaaaaa aaaaaagaaa gagaaggaaa aaatagtatc cattagagct atgggtatac 2400
catcacatcg atagtaaaaa gagagtcaaa agtcacacca ggacacaact agaacaaaaa 2460
aagggggcg cccccacttt gtaggccctc taaccggggg aatttttggg gcgggtacct 2520
ctgggggaac caattttccc caaaggcgcg cgtttttaag cctaagggga acaaaggggg 2580
aaaatattcc cccgggggga aattttttcc ccgcccacaa ttccccacaa ttaaggggg 2640
gggggggcta aaagttaaac acccgagggg ggcccatgg ggggggggaa ccccccttt 2700
tattgggtgg ggcacacgtg ccccgtttcc cccggggggg gaccctttt ttaccactat 2760
at 2762

<210> 7

<211> 223

<212> PRT

<213> Homo sapiens

<220> -

<223> 3577857

Met	Lys	Leu	His	Tyr	Val	Ala	Val	Leu	Thr	Leu	Ala	Ile	Leu	Met	
1				5					10					15	
Phe	Leu	Thr	Trp	Leu	Pro	Glu	Ser	Leu	Ser	Cys	Asn	Lys	Ala	Leu	
				20					25					30	
Cys	Ala	Ser	Asp	Val	Ser	Lys	Cys	Leu	Ile	Gln	Glu	Leu	Cys	Gln	
				35					40					45	
Cys	Arg	Pro	Gly	Glu	Gly	Asn	Cys	Ser	Cys	Cys	Lys	Glu	Cys	Met	
				50					55					60	
Leu	Cys	Leu	Gly	Ala	Leu	Trp	Asp	Glu	Cys	Cys	Asp	Cys	Val	Gly	
				65					70					75	
Met	Cys	Asn	Pro	Arg	Asn	Tyr	Ser	Asp	Thr	Pro	Pro	Thr	Ser	Lys	
				80					85					90	
Ser	Thr	Val	Glu	Glu	Leu	His	Glu	Pro	Ile	Pro	Ser	Leu	Phe	Arg	
				95					100					105	
Ala	Leu	Thr	Glu	Gly	Asp	Thr	Gln	Leu	Asn	Trp	Asn	Ile	Val	Ser	
				110					115					120	
Phe	Pro	Val	Ala	Glu	Glu	Leu	Ser	His	His	Glu	Asn	Leu	Val	Ser	
				125					130					135	
Phe	Leu	Glu	Thr	Val	Asn	Gln	Pro	His	His	Gln	Asn	Val	Ser	Val	
				140					145					150	
Pro	Ser	Asn	Asn	Val	His	Ala	Pro	Tyr	Ser	Ser	Asp	Lys	Glu	His	
				155					160					165	
Met	Cys	Thr	Val	Val	Tyr	Phe	Asp	Asp	Cys	Met	Ser	Ile	His	Gln	
				170					175					180	
Cys	Lys	Ile	Ser	Cys	Glu	Ser	Met	Gly	Ala	Ser	Lys	Tyr	Arg	Trp	
				185					190					195	
Phe	His	Asn	Ala	Cys	Cys	Glu	Cys	Ile	Gly	Pro	Glu	Cys	Ile	Asp	
				200					205					210	
Tyr	Gly	Ser	Lys	Thr	Val	Lys	Cys	Met	Asn	Cys	Met	Phe			
				215					220						

```
<210> 8
<211> 2204
<212> DNA
<213> Homo sapiens
```

<220> -
<223> 3577857

gcgggaggcg	cggcctggcc	tgcactcaa	agccgcccga	gcgcgccccg	ggctcggccg	60
accgcggcggg	gatctagggg	tgggcgactt	cgcgggaccg	tggcgcatgt	ttcctgggag	120
ttactgatca	tcttctttga	agaaacatga	agttacacta	tgttgyctgt	cttactctag	180
ccactcgtac	gttccctaga	tggcttcacg	aatcactgag	ctgtaaccaa	gcactctgtg	240
ctagtgatgt	gagcaaatgc	ctcattcagg	agctctgccg	gtgcggggcg	ggagaaggca	300
attgtctctg	ctgtaaggag	tgcatgctgt	gtcttggggc	ccttggggac	gagtgtgtgtg	360
actgtgttgg	tatgtgtaat	cctcgaaatt	atagtgaac	acctccaact	tcaaagagca	420
cagtggagga	gctcgatgaa	ccgatccctt	ctctcttccg	ggcactcaca	gaaggagata	480
ctcagttgaa	tgggaacatc	gtttctttcc	ctgttgccga	agaactttca	catcatgaga	540
atctggtttc	atttttagaa	actgtgaacc	agccacacca	ccagaatgtg	tctgtcccca	600
gcaataatgt	tcacgcgcct	tattccagtg	acaaagaaca	catgtgtact	gtggtttatt	660
ttgatgactg	catgtccata	catcagtgta	aaatatcctg	tgagttccat	ggagcatcca	720
aatatcgctg	gtttcataat	gcctgtctcg	agtqcatatg	tccaqaatgt	attgactatg	780

```

gtagtaaaac tgtcaaagt atgaactgca tgttttaag aagacaaatg caaaccaaag 840
caacttagta aaataatagg tataaaaagt tattctgtaa gtctgttggt tgtatcttgt 900
atcagaatcc cagtaagtta agttgtaaa agtttggaat aagtttcttt taaaaatatg 960
acatagccag tgatgtgttt aattatataa ctgttcttac tgattttatt gccccctagc 1020
aataagccct ttcctttgaa tacatgtaca actttgggtca tatgagaagc aggtgcgag 1080
agaattcctt gaaagatctg aggtttttta cataaagtct gatgtgggtt tcctctagca 1140
ttccaaaagg tttttgcttt gaaagtgtta gcagaagcat gttgatgtga attatgattt 1200
cttcattgtc tactgttagc acactgagtt tttatagttg cacatcattc ctcattgtgc 1260
cttgttttat ccattttata aatagagtag atatttgata taccactctg ataactcata 1320
taaaaaatc atcataaaaa gcttaatttc atccctttta tgttggtttt aaaaggtaaa 1380
tgcttaccat attttataat tgagaactct tacatagtag aatccattct ataatacatg 1440
tggtgacaaa gctttagaga aagtttccta ttctcttcca tttccctgc ccaaagtgtc 1500
gacataggca gtgatgaaga atctttacca agattttcag ggtgtacctt tgaaattgct 1560
ttaaatgcac tgctgggtga aataattagc aagcaaaagc gtttctgtga cttcaggtag 1620
cagcttaaa agcactaggg atggggaacg aatgccaat cagactccac ctgagcacc 1680
aggaaacagc ttgtaccctg gttaggaaat ggtgttctg aaaggggagg ctgagccagt 1740
gagagactga acttgtgcag ccttagccaa gacaaagcag tgtttttcag cagacggctg 1800
atgggacagg aattgaagaa gagaattgac tcgtatgaac aggacagggg gaaaatgctg 1860
ggaattataa tgggaaacaa aactatctat gttcatattt tgtaatatatt catttgtaa 1920
gtttatatct ggatataatg ttctttttaa acaagtataa tcatatcgtc ggagggttaa 1980
attatgaaat tttagaatct ctattcaaga tgatgttcac tccaaatata ctacagaatt 2040
tagtcaacat tttatataat gtttcaataa atgtttcttt caataaagat actatgtgcc 2100
cttcattagta ataaaatctc aatcttaaag catgaatcta aaacataaat atttatatta 2160
cagactcaat tttaaaaaat tactaattgc tctagtgcatt attt 2204

```

<210> 9

<211> 441

<212> PRT

<213> Mus musculus

<220> -

<223> g567206

<400> 9

```

Met Ala Leu Pro Ser Asn Phe Leu Leu Gly Val Cys Cys Phe Ala
  1           5           10          15
Trp Leu Cys Phe Leu Ser Ser Leu Ser Ser Gln Ala Ser Thr Glu
          20          25          30
Glu Ser Gln Ser Gly Ala Ser Glu Asn Val Glu Ser Glu Ala Asp
          35          40          45
Pro Trp Ser Leu Leu Leu Pro Val Asp Gly Thr Asp Arg Ser Gly
          50          55          60
Leu Leu Pro Pro Leu Phe Lys Val Leu Ser Asp Arg Arg Gly Glu
          65          70          75
Thr Pro Lys Leu Gln Pro Asp Ser Arg Ala Leu Tyr Tyr Met Lys
          80          85          90
Lys Leu Tyr Lys Thr Tyr Ala Thr Lys Glu Gly Val Pro Lys Pro
          95          100         105
Ser Arg Ser His Leu Tyr Asn Thr Val Arg Leu Phe Ser Pro Cys
          110         115         120
Ala Gln Gln Glu Gln Ala Pro Ser Asn Gln Val Thr Gly Pro Leu
          125         130         135
Pro Met Val Asp Leu Leu Phe Asn Leu Asp Arg Val Thr Ala Met
          140         145         150
Glu His Leu Leu Lys Ser Val Leu Leu Tyr Thr Leu Asn Asn Ser
          155         160         165

```


Ala	Ser	Ser	Ser	Ser	Thr	Val	Thr	Cys	Met	Cys	Asp	Leu	Val	Val
				170					175					180
Lys	Glu	Ala	Met	Ser	Ser	Gly	Arg	Ala	Pro	Pro	Arg	Ala	Pro	Tyr
				185					190					195
Ser	Phe	Thr	Leu	Lys	Lys	His	Arg	Trp	Ile	Glu	Ile	Asp	Val	Thr
				200					205					210
Ser	Leu	Leu	Gln	Pro	Leu	Val	Thr	Ser	Ser	Glu	Arg	Ser	Ile	His
				215					220					225
Leu	Ser	Val	Asn	Phe	Thr	Cys	Thr	Lys	Asp	Gln	Val	Pro	Glu	Asp
				230					235					240
Gly	Val	Phe	Ser	Met	Pro	Leu	Ser	Val	Pro	Pro	Ser	Leu	Ile	Leu
				245					250					255
Tyr	Leu	Asn	Asp	Thr	Ser	Thr	Gln	Ala	Tyr	His	Ser	Trp	Gln	Ser
				260					265					270
Leu	Gln	Ser	Thr	Trp	Arg	Pro	Leu	Gln	His	Pro	Gly	Gln	Ala	Gly
				275					280					285
Val	Ala	Ala	Arg	Pro	Val	Lys	Glu	Glu	Ala	Thr	Glu	Val	Glu	Arg
				290					295					300
Ser	Pro	Arg	Arg	Arg	Arg	Gly	Gln	Lys	Ala	Ile	Arg	Ser	Glu	Ala
				305					310					315
Lys	Gly	Pro	Leu	Leu	Thr	Ala	Ser	Phe	Asn	Leu	Ser	Glu	Tyr	Phe
				320					325					330
Lys	Gln	Phe	Leu	Phe	Pro	Gln	Asn	Glu	Cys	Glu	Leu	His	Asp	Phe
				335					340					345
Arg	Leu	Ser	Phe	Ser	Gln	Leu	Lys	Trp	Asp	Asn	Trp	Ile	Val	Ala
				350					355					360
Pro	His	Arg	Tyr	Asn	Pro	Arg	Tyr	Cys	Lys	Gly	Asp	Cys	Pro	Arg
				365					370					375
Ala	Val	Arg	His	Arg	Tyr	Gly	Ser	Pro	Val	His	Thr	Met	Val	Gln
				380					385					390
Asn	Ile	Ile	Tyr	Glu	Lys	Leu	Asp	Pro	Ser	Val	Pro	Arg	Pro	Ser
				395					400					405
Cys	Val	Pro	Gly	Lys	Tyr	Ser	Pro	Leu	Ser	Val	Leu	Thr	Ile	Glu
				410					415					420
Pro	Asp	Gly	Ser	Ile	Ala	Tyr	Lys	Glu	Tyr	Glu	Asp	Met	Ile	Ala
				425					430					435
Thr	Arg	Cys	Thr	Cys	Arg									
				440										

<210> 10

<211> 240

<212> PRT

<213> Homo sapiens

<220> -

<223> g598956

<400> 10

Met	Ser	Arg	Ser	Asn	Arg	Gln	Lys	Glu	Tyr	Lys	Cys	Gly	Asp	Leu
1				5					10					15
Val	Phe	Ala	Lys	Met	Lys	Gly	Tyr	Pro	His	Trp	Pro	Ala	Arg	Ile
				20					25					30
Asp	Glu	Met	Pro	Glu	Ala	Ala	Val	Lys	Ser	Thr	Ala	Asn	Lys	Tyr
				35					40					45

Gln	Val	Phe	Phe	Phe	Gly	Thr	His	Glu	Thr	Ala	Phe	Leu	Gly	Pro
				50					55					60
Lys	Asp	Leu	Phe	Pro	Tyr	Glu	Glu	Ser	Lys	Glu	Lys	Phe	Gly	Lys
				65					70					75
Pro	Asn	Lys	Arg	Lys	Gly	Phe	Ser	Glu	Gly	Leu	Trp	Glu	Ile	Glu
				80					85					90
Asn	Asn	Pro	Thr	Val	Lys	Ala	Ser	Gly	Tyr	Gln	Ser	Ser	Gln	Lys
				95					100					105
Lys	Ser	Cys	Val	Glu	Glu	Pro	Glu	Pro	Glu	Pro	Glu	Ala	Ala	Glu
				110					115					120
Gly	Asp	Gly	Asp	Lys	Lys	Gly	Asn	Ala	Glu	Gly	Ser	Ser	Asp	Glu
				125					130					135
Glu	Gly	Lys	Leu	Val	Ile	Asp	Glu	Pro	Ala	Lys	Glu	Lys	Asn	Glu
				140					145					150
Lys	Gly	Ala	Leu	Lys	Arg	Arg	Ala	Gly	Asp	Leu	Leu	Glu	Asp	Ser
				155					160					165
Pro	Lys	Arg	Pro	Lys	Glu	Ala	Glu	Asn	Pro	Glu	Gly	Glu	Glu	Lys
				170					175					180
Glu	Ala	Ala	Thr	Leu	Glu	Val	Glu	Arg	Pro	Leu	Pro	Met	Glu	Val
				185					190					195
Glu	Lys	Asn	Ser	Thr	Pro	Ser	Glu	Pro	Gly	Ser	Gly	Arg	Gly	Pro
				200					205					210
Pro	Gln	Glu	Glu	Glu	Glu	Glu	Glu	Asp	Glu	Glu	Glu	Glu	Ala	Thr
				215					220					225
Lys	Glu	Asp	Ala	Glu	Ala	Pro	Gly	Ile	Arg	Asp	His	Glu	Ser	Leu
				230					235					240

<210> 11

<211> 714

<212> PRT

<213> *Saccharomyces cerevisiae*

<220> -

<223> g500832

<400> 11

Met	Ser	His	Leu	Ile	Thr	Leu	Ala	Thr	Cys	Asn	Leu	Asn	Gln	Trp
1				5					10					15
Ala	Leu	Asp	Phe	Glu	Gly	Asn	Arg	Asp	Arg	Ile	Leu	Gln	Ser	Ile
				20					25					30
Lys	Ile	Ala	Lys	Glu	Arg	Gly	Ala	Arg	Leu	Arg	Val	Gly	Pro	Glu
				35					40					45
Leu	Glu	Ile	Thr	Gly	Tyr	Gly	Cys	Leu	Asp	His	Phe	Leu	Glu	Asn
				50					55					60
Asp	Val	Cys	Leu	His	Ser	Trp	Glu	Met	Tyr	Ala	Gln	Ile	Ile	Lys
				65					70					75
Asn	Lys	Glu	Thr	His	Gly	Leu	Ile	Leu	Asp	Ile	Gly	Met	Pro	Val
				80					85					90
Leu	His	Lys	Asn	Val	Arg	Tyr	Asn	Cys	Arg	Leu	Leu	Ser	Leu	Asp
				95					100					105
Gly	Glu	Ile	Leu	Phe	Ile	Arg	Pro	Lys	Ile	Trp	Leu	Ala	Asn	Asp
				110					115					120
Gly	Asn	Tyr	Arg	Glu	Met	Arg	Phe	Phe	Thr	Pro	Trp	Met	Lys	Pro
				125					130					135

Gly Val Val Glu Asp Phe Ile Leu Pro	Pro Glu Ile Gln Lys Val	140	145	150
Thr Gly Gln Arg Leu Val Pro Phe Gly	Asp Ala Val Ile Asn Ser	155	160	165
Leu Asp Thr Cys Ile Gly Thr Glu Thr	Cys Glu Glu Leu Phe Thr	170	175	180
Pro Gln Ser Pro His Ile Ala Met Ser	Leu Asp Gly Val Glu Ile	185	190	195
Met Thr Asn Ser Ser Gly Ser His His	Glu Leu Arg Lys Leu Asn	200	205	210
Lys Arg Leu Asp Leu Ile Leu Asn Ala	Thr Lys Arg Cys Gly Gly	215	220	225
Val Tyr Leu Tyr Ala Asn Gln Arg Gly	Cys Asp Gly Asp Arg Leu	230	235	240
Tyr Tyr Asp Gly Cys Ala Leu Ile Ala	Ile Asn Gly Thr Ile Val	245	250	255
Ala Gln Gly Ser Gln Phe Ser Leu Asp	Asp Val Glu Val Val Thr	260	265	270
Ala Thr Val Asp Leu Glu Glu Val Arg	Ser Tyr Arg Ala Ala Val	275	280	285
Met Ser Arg Gly Leu Gln Ala Ser Leu	Ala Glu Ile Lys Phe Lys	290	295	300
Arg Ile Asp Ile Pro Val Glu Leu Ala	Leu Met Thr Ser Arg Phe	305	310	315
Asp Pro Thr Val Cys Pro Thr Lys Val	Arg Glu Pro Phe Tyr His	320	325	330
Ser Pro Glu Glu Glu Ile Ala Leu Gly	Pro Ala Cys Trp Met Trp	335	340	345
Asp Tyr Leu Arg Arg Cys Asn Gly Thr	Gly Phe Phe Leu Pro Leu	350	355	360
Ser Gly Gly Ile Asp Ser Cys Ala Thr	Ala Met Ile Val His Ser	365	370	375
Met Cys Arg Leu Val Thr Asp Ala Ala	Gln Asn Gly Asn Glu Gln	380	385	390
Val Ile Lys Asp Val Arg Lys Ile Thr	Arg Ser Gly Asp Asp Trp	395	400	405
Ile Pro Asp Ser Pro Gln Asp Leu Ala	Ser Lys Ile Phe His Ser	410	415	420
Cys Phe Met Gly Thr Glu Asn Ser Ser	Lys Glu Thr Arg Asn Arg	425	430	435
Ala Lys Asp Leu Ser Asn Ala Ile Gly	Ser Tyr His Val Asp Leu	440	445	450
Lys Met Asp Ser Leu Val Ser Ser Val	Val Ser Leu Phe Glu Val	455	460	465
Ala Thr Gly Lys Lys Pro Ile Tyr Lys	Ile Phe Gly Gly Ser Gln	470	475	480
Ile Glu Asn Leu Ala Leu Gln Asn Ile	Gln Ala Arg Leu Arg Met	485	490	495
Val Leu Ser Tyr Leu Phe Ala Gln Leu	Leu Pro Trp Val Arg Gly	500	505	510
Ile Pro Asn Ser Gly Gly Leu Leu Val	Leu Gly Ser Ala Asn Val	515	520	525
Asp Glu Cys Leu Arg Gly Tyr Leu Thr	Lys Tyr Asp Cys Ser Ser	530	535	540
Ala Asp Ile Asn Pro Ile Gly Gly Ile	Ser Lys Thr Asp Leu Lys	545	550	555
Arg Phe Ile Ala Tyr Ala Ser Lys Gln	Tyr Asn Met Pro Ile Leu			

560	565	570
Asn Asp Phe Leu Asn Ala Thr Pro Thr	Ala Glu Leu Glu Pro Met	
575	580	585
Thr Lys Asp Tyr Val Gln Ser Asp Glu	Ile Asp Met Gly Met Thr	
590	595	600
Tyr Glu Glu Leu Gly Val Phe Gly Tyr	Leu Arg Lys Val Glu Lys	
605	610	615
Cys Gly Pro Tyr Ser Met Phe Leu Lys	Leu Leu His Gln Trp Ser	
620	625	630
Pro Lys Leu Thr Pro Arg Gln Ile Ser	Glu Lys Val Lys Arg Phe	
635	640	645
Phe Phe Phe Tyr Ala Ile Asn Arg His	Lys Gln Thr Val Leu Thr	
650	655	660
Pro Ser Tyr His Ala Glu Gln Tyr Ser	Pro Glu Asp Asn Arg Phe	
665	670	675
Asp Leu Arg Pro Phe Leu Ile Asn Pro	Arg Phe Pro Trp Ala Ser	
680	685	690
Arg Lys Ile Asp Glu Val Val Glu Gln	Cys Glu Ala His Lys Gly	
695	700	705
Ser Thr Leu Asp Ile Met Ser Ile Asp		
710		

<210> 12
 <211> 249
 <212> PRT
 <213> Drosophila melanogaster

<220> -
 <223> g529900

<400> 12

Met Gln Leu Leu Cys Tyr Phe Val Ile Leu Phe Val Gly Ile Ala	
1 5 10 15	
Pro Trp Ser Ser Leu Ala Asn Asp Asp Gly Cys Asn Glu Val Val	
20 25 30	
Cys Gly Ser Val Val Ser Lys Cys Leu Ile Thr Gln Ser Cys Gln	
35 40 45	
Cys Lys Leu Asn Asp Cys His Cys Cys Lys Asp Cys Leu Asn Cys	
50 55 60	
Leu Gly Glu Leu Tyr Ile Glu Cys Cys Gly Cys Leu Asp Met Cys	
65 70 75	
Pro Lys His Lys Asp Val Leu Pro Ser Leu Thr Pro Arg Ser Glu	
80 85 90	
Ile Gly Asp Ile Glu Gly Val Pro Glu Leu Phe Asp Thr Leu Thr	
95 100 105	
Ala Glu Asp Asp Glu Gly Trp Ser Thr Ile Arg Phe Ser Met Arg	
110 115 120	
Ala Gly Phe Lys Gln Arg Val Gln Gly Gly Ala Ser Gly Asp Ala	
125 130 135	
Gly Asn Gly Asn Gly Asn Gly Asn Ala Gly Ser Ala Gly Val Thr	
140 145 150	
Leu Cys Thr Val Ile Tyr Val Asn Ser Cys Ile Arg Ala Asn Lys	
155 160 165	
Cys Arg Gln Gln Cys Glu Ser Met Gly Ala Ser Ser Tyr Arg Trp	

	170		175		180
Phe His Asp Gly Cys Cys Glu Cys Val	Gly Glu Asn Cys Leu Asn				
185	190				195
Tyr Gly Ile Asn Glu Ser Arg Cys Arg	Gly Cys Pro Glu Asp Gln				
200	205				210
Asp Gln Leu Leu Thr Ala Asp Thr Val	Pro Ala Glu Ala Glu Gln				
215	220				225
Asp Leu Glu Arg Phe Phe Gly Asn Glu	Glu Ile Glu Asp Glu Trp				
230	235				240
Gly Tyr Gly Glu Glu Asp Glu Phe Ser					
245					

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/12, C07K 14/475, A61K 38/18, C07K 16/22, C12Q 1/68		A3	(11) International Publication Number: WO 99/38976
			(43) International Publication Date: 5 August 1999 (05.08.99)
(21) International Application Number: PCT/US99/00654		TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US).	
(22) International Filing Date: 12 January 1999 (12.01.99)		(74) Agents: BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).	
(30) Priority Data: 09/015,412 29 January 1998 (29.01.98) US		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application US 09/015,412 (CIP) Filed on 29 January 1998 (29.01.98)		Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
(71) Applicant (for all designated States except US): INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).		(88) Date of publication of the international search report: 25 November 1999 (25.11.99)	
(72) Inventors; and (75) Inventors/Applicants (for US only): HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #12, Mountain View, CA 94040 (US). LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). CORLEY, Neil, C. [US/US]; 1240 Dale Avenue #30, Mountain View, CA 94040 (US). BAUGHN, Mariah [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). GUEGLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US).			
(54) Title: HUMAN GROWTH FACTOR HOMOLOGS			
(57) Abstract <p>The invention provides human growth factor homologs (HGFH) and polynucleotides which identify and encode HGFH. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for treating or preventing disorders associated with expression of HGFH.</p>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 99/00654

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/12 C07K14/475 A61K38/18 C07K16/22 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C07K A61K C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 94 15966 A (UNIV JOHNS HOPKINS MED ; LEE SE JIN (US)) 21 July 1994 see whole document, particularly seq.26 (human GDF-9 protein seq). ---	1-9, 11-21
Y	DATABASE EMBL - EMBEST1 Entry/Acc.no. AA604716, 30 July 1997 STRAUSBERG, R.: "no76e01.s1 NCI CGAP AA1 homo sapiens cDNA clone IMAGE:1112760 similar to SW:GDF9 MOUSE Q07105 GROWTH/DIFFERENTIATION FACTOR 9 PRECURSOR." XP002105030 see the whole document --- -/--	1-13, 18-21

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "Z" document member of the same patent family

Date of the actual completion of the international search

7 June 1999

Date of mailing of the international search report

01.10.1999

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Smalt, R

INTERNATIONAL SEARCH REPORT

Internat. Application No.
PCT/US 99/00654

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	MCPHERRON A C ET AL: "GDF-3 AND GDF-9: TWO NEW MEMBERS OF THE TRANSFORMING GROWTH FACTOR -B SUPERFAMILY CONTAINING A NOVEL PATTERN OF CYSTEINES" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 268, no. 5, 15 February 1993, pages 3444-3449, XP000348932 see the whole document ---	1-13, 18-21
A	WO 96 39485 A (HUMAN GENOME SCIENCES INC ;ROSEN CRAIG A (US); KUNSCH CHARLES A (U) 12 December 1996 see the whole document ---	
A	WO 97 34998 A (HUMAN GENOME SCIENCES INC ;LI YI (US); LI HAODONG (US); SU KUI (US) 25 September 1997 see the whole document ---	
P,X	DATABASE EMBL - EMHUM1 Entry/Acc.no. AC004500, 6 April 1998 KIMMERLY, W. ET AL.: "Homo sapiens chromosome 5, P1 clone 1076B9 (LBNL H14), complete sequence." XP002105031 nt 39423-42780 -----	3-10

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 99/00654

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

see remark
2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-21 (all partially)

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Remark:

Although claims 18 and 19 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Further defect(s) under Article 17(2)(a):

Claims 16 and 17, referring to agonists and antagonists, respectively could not be searched to completion due to insufficient characterization of said compounds in the description.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

1. Claims: 1-21, all partially

HGFH-1 polypeptide and variants thereof, polynucleotides encoding them, polynucleotides which hybridize thereto, fragments thereof, expression vector comprising said polynucleotide, host cell transformed with said vector, method of producing said protein using said host, antibody and (ant)agonists against said protein, pharmaceutical composition containing said protein or said antagonist, and a method of detecting said nucleic acid through hybridization and/or PCR.

2. Claims: 1-21, all partially

HGFH-2 polypeptide and variants thereof, polynucleotides encoding them, polynucleotides which hybridize thereto, fragments thereof, expression vector comprising said polynucleotide, host cell transformed with said vector, method of producing said protein using said host, antibody and (ant)agonists against said protein, pharmaceutical composition containing said protein or said antagonist, and a method of detecting said nucleic acid through hybridization and/or PCR.

3. Claims: 1-21, all partially

HGFH-3 polypeptide and variants thereof, polynucleotides encoding them, polynucleotides which hybridize thereto, fragments thereof, expression vector comprising said polynucleotide, host cell transformed with said vector, method of producing said protein using said host, antibody and (ant)agonists against said protein, pharmaceutical composition containing said protein or said antagonist, and a method of detecting said nucleic acid through hybridization and/or PCR.

4. Claims: 1-21, all partially

HGFH-4 polypeptide and variants thereof, polynucleotides encoding them, polynucleotides which hybridize thereto, fragments thereof, expression vector comprising said polynucleotide, host cell transformed with said vector, method of producing said protein using said host, antibody and (ant)agonists against said protein, pharmaceutical composition containing said protein or said antagonist, and a method of detecting said nucleic acid through hybridization and/or PCR.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 99/00654

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9415966 A	21-07-1994	CA 2153653 A EP 0678101 A JP 9508001 T US 5821056 A	21-07-1994 25-10-1995 19-08-1997 13-10-1998
WO 9639485 A	12-12-1996	CA 2223733 A AU 2692295 A EP 0833892 A	12-12-1996 24-12-1996 08-04-1998
WO 9734998 A	25-09-1997	AU 5369296 A EP 0907719 A	10-10-1997 14-04-1999

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

1. Claims: 1-21, all partially

HGFH-1 polypeptide and variants thereof, polynucleotides encoding them, polynucleotides which hybridize thereto, fragments thereof, expression vector comprising said polynucleotide, host cell transformed with said vector, method of producing said protein using said host, antibody and (ant)agonists against said protein, pharmaceutical composition containing said protein or said antagonist, and a method of detecting said nucleic acid through hybridization and/or PCR.

2. Claims: 1-21, all partially

HGFH-2 polypeptide and variants thereof, polynucleotides encoding them, polynucleotides which hybridize thereto, fragments thereof, expression vector comprising said polynucleotide, host cell transformed with said vector, method of producing said protein using said host, antibody and (ant)agonists against said protein, pharmaceutical composition containing said protein or said antagonist, and a method of detecting said nucleic acid through hybridization and/or PCR.

3. Claims: 1-21, all partially

HGFH-3 polypeptide and variants thereof, polynucleotides encoding them, polynucleotides which hybridize thereto, fragments thereof, expression vector comprising said polynucleotide, host cell transformed with said vector, method of producing said protein using said host, antibody and (ant)agonists against said protein, pharmaceutical composition containing said protein or said antagonist, and a method of detecting said nucleic acid through hybridization and/or PCR.

4. Claims: 1-21, all partially

HGFH-4 polypeptide and variants thereof, polynucleotides encoding them, polynucleotides which hybridize thereto, fragments thereof, expression vector comprising said polynucleotide, host cell transformed with said vector, method of producing said protein using said host, antibody and (ant)agonists against said protein, pharmaceutical composition containing said protein or said antagonist, and a method of detecting said nucleic acid through hybridization and/or PCR.